Some Formal Solutions in Side-channel Cryptanalysis

An Introduction

Fabrice J.P.R. Pautot
e-mail to: fabrice.pautot@laposte.net
22/12/08

Abstract: We propose to revisit Side-channel Cryptanalysis from the point of view, for instance, of C. E. Shannon: *The calculation of a posteriori probabilities is the generalized problem of cryptanalysis*. So, our goal will be to provide analytic formulae for the marginal posterior probability mass functions *for* the targets of those attacks. Since we are concerned with the probabilities of single and perfectly determined cases, we need above all to place ourselves in a probabilistic system enjoying an epistemic "interpretation". We select *Probability as Logic*, the most suitable system for our purpose. With this powerful and flexible system at hand, we first solve two independent problems for known, non-chosen messages: the determination of side-channel *leakage times* (generalized for high-order attacks) and the determination of the target, given those *leakage times*. The first problem belongs to Hypotheses Testing Theory and admits a formal solution in terms of Bayes Factors in the parametric framework. The calculation of those factors requires marginalizing over all possible values of the target, so that this new procedure has no equivalent in frequentist Statistics and we indicate how it could be proved to outperform previous procedures more and more, as the target space size increases. We present preliminary experimental results and give some clues on how to extend this solution to the nonparametric framework. The second problem is a classical Parameter Estimation problem with many hyperparameters. It also admits a unique *maximum a posteriori* solution under 0-1 loss function within Decision Theory. When it is not possible to solve both problems independently, we must solve them simultaneously in order to get general solutions for Side-channel Cryptanalysis on symmetric block ciphers, at least. Taking benefit of the duality between Hypotheses Testing and Parameter Estimation in our system of inference, we transform the determination of the *generalized leakage times* into a parameter estimation problem, in order to fall back into a global parameter estimation problem. Generally speaking, it appears that (marginal) side-channel parametric leakage models are in fact averages between attack and "non-attack" models and, more generally between many conditional models, so that likelihoods can not be frequency sampling distributions. Then, we give the marginal posterior probability mass function for the targets of the most general known-messages attacks: "correlation" attacks, template attacks, high-order attacks, multi-*decision functions* attacks, multi-*attack models* attacks and multi-*"non-attack" models* attacks. Essentially, it remains to explain how to assign joint prior and discrete direct probability distributions by logical inspection, to extent this approach to the nonparametric framework and other cryptographic primitives, to deal with analytic, symbolic, numerical and computational implementation issues and especially to derive formal adaptive chosen-messages attacks.

Keywords: (Side-channel) Cryptanalysis, Differential Power Analysis (DPA), Template Attacks, High-order Attacks, Statistical Inference, Plausible Reasoning, Probability as Logic, Principle of Maximum Entropy, Hypotheses Testing, Bayes Factors, Parameter Estimation, Maximum *a posteriori* Estimator, Formal Methods, Security Proofs.

Introduction – The Need for Some Probability Theory

We propose to revisit Side-channel Cryptanalysis, as introduced by Kocher [47], by restarting from scratch with a natural, intuitive, appealing and elementary principle or guide:

Principle: *In order to perform a (side-channel) attack, it should be necessary and sufficient to compute the probability distribution for its target and to take, as our best guess, the value in the target space having highest probability.*

This Principle seems to be almost as old as the World itself. For instance, it is already stated as it is, in a non-cryptographic context, by Mister Aristotle himself [3]:

Find out what you think is a good life and consider the probabilities of your possible actions to achieve this. Then follow the course of action which with highest probability results in a good life.

We find also much stronger variations on our Principle. According to J. C. Maxwell [54], the great master of EMA attacks [67][72][80]:

The actual science of logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which (fortunately) we have to reason on. Therefore, the true logic for this world is the calculus of probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind.

More recently, it was recalled by C. E. Shannon in the cryptographic context [78]:

The calculation of a posteriori probabilities is the generalized problem of cryptanalysis.

Here we already find something intriguing. Recently, Gierlichs, Batina and Tuyls [30] and Aumônier [5] proposed *information-theoretic differential side-channel attacks* based on empirical Shannon *Mutual Informations* (i.e. *Mutual Information Analysis* or MIA). This approach is interesting as it is nonparametric: it does not require any assumption on the underlying side-channel leakage model(s) by contrast, for instance, to more or less implicit linearity and Gaussian assumptions for linear correlation coefficients (i.e. CPA [21]). But, astonishingly, Shannon does not tell us to compute *his* empirical mutual informations but *a posteriori* probabilities instead! Clearly, there is something to understand. Later, we will provide some clues on how to make the asymptotic link between MIA and the present approach explicit.

So, the purpose of this paper is to provide analytic formulae for the marginal posterior probability mass function for the targets of standard, generic side-channel attacks, at least on symmetric block ciphers.

The problem is that, as soon as we try to follow this Principle, we fall into trouble and draw quite deep and disturbing conclusions. Indeed, if we can ever conceive that the targets of those side-channel attacks, typically *subkeys*, can admit probability distributions in some system of probability to be determined, then we should certainly acknowledge that this system has essentially nothing to do with a very popular and common mathematical abstraction: random variables. Indeed, since the subkeys are logical constants that are perfectly determined, fixed and stored once and for all in our embedded devices, there is absolutely nothing “random” and nothing “variable” in our story, but the opposite. The problem is that the theory of probability that you and me were taught at school deals *essentially* with random variables [48]: just check Chapter 2 of your favourite probabilistic textbook. In the same way, this theory is *supposed* to describe *mass random phenomena* or events [48][60], while we are here dealing with Boolean propositions such as

‘the (sub)key k stored in this microcontroller has value k_0’

that do not really correspond to any (mass) (random) event.

Similarly, we were also taught that probabilities are in fact defined as or from relative frequencies “in the long run” [48][60], in an extremely complex way, while there is again no frequency at all in our story but just THE key to be determined.

So, there is clearly a deep semantic gap between the theory of “probability” we know well and what we need in order to be prepared to follow Shannon way to (Side-channel) Cryptanalysis. But does it really mean that we cannot use and “interpret” the abstract, formal, mathematical system that we are familiar with in a radically different, completely opposite context from the one for which it was originally designed? Do we really need a brand new formal, axiomatic system of probability for our purpose? Unfortunately, the answer is “yes, we absolutely need a new system.”. Even, the confidentiality of such suitable systems is probably the main reason why, to the best of our knowledge, nobody has proposed to follow the Principle and Shannon way so far, at least in Side-channel Cryptanalysis.

This is not the purpose of the present paper to explain why, but a careful logical inspection that we may provide in details in a companion paper shows us that we definitely need a system of probability:

- that deals with arbitrary Boolean propositions, not only (mass) (random) events [48][60].
- that does not apply only to repetitive events but also to single cases [12].
- that does not *essentially* deal with random variables.
- in which probabilities are not defined as or from relative frequencies [14][48][60].
- whose mathematical axioms are not justified empirically [48] or by equally likely cases [77].
- that is not hypothetico-deductive [40].

Many such systems or theories of probability exist in the literature. For instance, according to another proponent of the Principle, the great cryptanalyst I.J. Good (assistant of Alan Turing during WWII) [33], a least 46656 such theories exist and perhaps infinitely many of them!

However we must make a choice. In this paper we will use and apply a theory known as *Plausibility Theory* [4][24][25][26][28][36][38][46][56][70][71][77] or *Probability as* (*Extended*) *Logic* [17][18][19][20][40]. It belongs to the family of *Objective* (rather *Intersubjective*) *Bayesian Theories* [7][40][43][74]. After all, we can say that *Probability as Logic* is just *Classical Probability Theory beyond the Principle of Insufficient Reason.*

The reader is urged to check Bretthorst [17][18] first for excellent tutorials and [20] for impressive applications in Spectrum Analysis. Jaynes *opus magnus* [40] is the absolute reference in this field. [39] is of historical importance.

From the historical point of view, first-class direct or indirect contributors to *Probability as Logic* include, in rough chronological order: Pascal [66], de Fermat, Jakob Bernoulli [8], Leibniz [53], de Moivre [61], Bayes, Laplace [51][52], Maxwell [54], Gibbs [29], Bertrand [9], Poincaré [68], Keynes [46], Borel [12][13][14], de Finetti [27], Jeffreys [43], Cox [24][25], Shannon [78], Pólya [69] and Jaynes [39][40][41][42]. Note that most of those eminent scientists were also competent physicists. This is certainly a good point, since we are definitely concerned with some physics here. Apart from Shannon and Good, we have been able to find one and only one weak connection between modern Cryptology and *Probability as Logic*: [34] by Rivest *et al.*

Let us briefly summarize the main concepts of *Probability as Logic* to make this paper fairly self-contained without bothering the reader too much with those theoretical preliminaries:

- You start with a new informal concept in Boolean Logic: when you do not have enough information to determine or to know if a Boolean proposition A is true or false, you say that it is more or less *plausible*. This concept of *plausibility* is clearly *relative*: a Boolean proposition A is plausible with respect to/given/conditionally upon another Boolean proposition I . I has many names in the literature, for instance the *corpus/group/system of knowledge* according to Keynes and Borel [46][12] or simply the *background information* according to Jaynes [40]. This is the logical conjunction of all Boolean statements true or assumed to be true and relevant for proposition A .

- You seek for the rules governing those *degrees of plausibility* (i.e. proto-probabilities), in symbols $\pi\left(A|I\right)$, in order to get *extensions of Boolean Logic under uncertainty*. Assuming that they can be represented by real numbers (a natural but strong assumption), you can prove, under some common-sense qualitative *desiderata* and some technical conditions, that for the sake of logical consistency those rules are necessarily, up to some isomorphism/convention

$$\begin{cases}\pi\left(A\wedge B|I\right)=\pi\left(A|B\wedge I\right)\pi\left(B|I\right)\\ \pi\left(\neg A|I\right)=1-\pi\left(A|I\right)\end{cases}$$

  that are nothing but the *Product Rule* and the *Sum Rule* of Probability Theory respectively. Those kind of results are known as *Cox-style Theorems* [4][24][25][26][28][36][38][77]. The result is essentially the same as de Finetti *Dutch Book Theorem/Bets Method* [12][27][40] but you do not need to introduce *utilities* and stick to probability.

- Because you try to describe *states of knowledge* instead of (mass) (random) phenomena [40][48][60], the fundamental concept of *independence* is no longer reduced to causal/stochastic independence but is the logical conjunction of causal independence and *logical independence* [40]. Two Boolean propositions are said to be logically independent if and only if the knowledge on one of them has no influence on the knowledge on the other and conversely, etc. As a consequence, unconditional independence, iidness and random variables disappear in *Probability as Logic*. In this way, we get a *predictive theory*. Unconditional iidness is replaced by *conditional iidness*, which is guaranteed, for instance, for infinitely and finitely discrete equivalent/exchangeable propositions by *Integral Representation Theorems*

[7][27][41][45]. See also Porta Mana [56][70][71] for an alternative interpretation of Integral Representations. As a result we have only prior, direct (i.e. likelihoods and parametric models) and posterior probability distributions, instead of random variables having prior and posterior probability distributions.

- The conversion of the *background information I* into prior and direct probability distributions by logical inspection still needs to be achieved. This is the very goal of *Probability as Logic*. For this purpose, a large toolbox is available. Depending on the situation, we can use tools such as: the good old *Principle of Insufficient Reason/Indifference* (Pascal [66], Jakob Bernoulli [8], Leibniz [53], Laplace [51][52], Poincaré [68], Keynes [46]) for equally likely cases, the *Method of Arbitrary Functions* (Poincaré [68], Borel [13]), the *Principle of Maximum Entropy* (Jaynes [39][40][42]), logical invariance under group transformation (Borel [13], Lhoste [55], Jeffreys [43], Jaynes [40]), *Reference Priors* (Bernardo [7]), *Marginalization Theory* (Jaynes [40]), *Coding Theory* (Rissanen [73]), *Information Geometry* (Rodriguez [75]), etc. In this paper we do not deal with the assignment of prior and direct probability distributions. Therefore we will keep both *background information I* and models in our symbolic notations.

It is absolutely essential to keep in mind that *Probability as Logic* describes only *states of knowledge* and not (mass) (random) events or phenomena themselves. This is the reason why we must and will talk about probabilities (i.e. plausibilities) *for* Boolean propositions and not "probabilities" (e.g. "relative frequencies") *of* events. The confusion between *epistemic* and *ontic* theories is called the *Mind Projection Fallacy* and is known to have drastic quantitative consequences in applications [40]. Let us illustrate briefly the gap between both kinds of theories on two important examples that are highly relevant for our purpose:

- **The Gaussian "assumption".** In ontic theories, it is *assumed* that some i.i.d. events *have* a Laplace-Gauss frequency sampling distribution. This assumption can be proved or rather justified by applying for instance *Central Limit-like Theorems* or, more generally, *Errors Theory* [68]. Subsequently, we just fall on other assumptions (e.g. large number of small symmetric errors, etc.) that we should in any case verify *a posteriori* on our experimental data in order to justify our working hypotheses and to apply what we were able to deduce from them (and, in many cases, it is easier to verify the Gaussian assumption itself.). However, in *Probability as Logic*, there is no Gaussian assumption or hypothesis. Laplace-Gauss distributions arise just as logical consequences of applications of the *Principle of Maximum Entropy* that tells us that, among all probability distributions satisfying some constraints (e.g. first and second moments for Laplace-Gauss distributions), we must select the one having maximum entropy, because, Shannon (differential) entropy being the unique measure of uncertainty in so many regards, it is the most uncertain [39][40][42]. Thus, in assigning and introducing epistemic Laplace-Gauss distributions within *Probability as Logic*, we just try to be as honest as possible. And trying to be honest is not an assumption: it is a duty. An immediate but naïve objection is the following: "If you are concerned only with uncertain, "honest" epistemic probability distributions, then you loose all contacts with physical, empirical phenomena. So, how do you explain for instance that so many observable phenomena seem to have Laplace-Gauss frequency sampling distributions???" And the answer is: "You know, this is precisely because we are physicists that we focus on epistemic distributions instead of empirical or counterfactual ones. We do observe maximum entropic epistemic probability distributions such as Laplace-Gauss in Mother Nature because it appears that they are exponentially more probable than any other." This is known as the *Entropy Concentration Phenomenon* [40][42]. Then, another remarkable phenomenon appears: basically, when using those constrained epistemic maximum entropic distributions, all frequency sampling distributions satisfying the same constraints lead to the same inferences at the end [19][40]. So, *Probability as Logic* actually makes frequency sampling distributions almost irrelevant, while frequentist Statistics [62][63] are, by definition, entirely based on them! For all those reasons, (multivariate) Laplace-Gauss distributions would also play a central role in Side-channel Cryptanalysis (as they already do… as assumptions) if our poor but expensive digital instruments would not provide us only with drastically discretized data (typically 8 bits precision.).

- **The conditional independence "assumption".** Depending on our *background information*, we can assign (conditionally) logically independent epistemic probability distributions for (the Boolean propositions corresponding to) events whose (unconditional) frequency sampling distributions are dependent. Conversely, we can assign (conditionally) logically dependent epistemic probability distributions for events whose (unconditional) frequency sampling distributions are independent. For instance, if we apply the *Principle of Maximum Entropy* to get the joint (conditional) multivariate direct probability distribution for the noise then, as we all know, the entropy is maximum for mutually (conditionally) logically independent marginal noise distributions [19]. Again, this should not be interpreted as a drastic, arbitrary assumption of mutual independence but, on the contrary, as making allowance for every possible correlation that could be present.

Now, let's try to apply *Probability as Logic* to Side-channel Cryptanalysis. In section I, we introduce our notations and terminology. In section II, we solve two generic sub-problems of interest and give some examples, some preliminary experimental results and some elements of a theoretical analysis of those new procedures. In section III, we merge both sub-problems in order to get fairly general and formal solutions to generic Side-channel Cryptanalysis.

I Side-channel Notations and Terminology

Finding good symbolic notations for Side-channel Cryptanalysis is definitely a very significant part of the full job! As we shall see, it is very easy to get lost in all those dimensions.

- A the *Attacker* (Eve)

- I^A the Keynes-Borel *corpus/system of knowledge* or *background information* of A

- '.' or X Boolean propositions, $\wedge$ the logical conjunction, $\neg$ the logical negation

- $p(X|I)$ the epistemic intersubjective plausibility/probability for Boolean proposition X conditional on/given *background information* I

- $*$ the convolution product

- $k = 0, K-1$ the possible values of the *subkey* to be determined

- k_0 the actual value of the subkey to be determined

- $i = 1, N$ the *encryptions*

- t the *time*

- $t_j, j = 1, J$ the *times* in the *attack window* (not necessarily contiguous)

- $t_j^n = \overset{n}{\underset{l=1}{\wedge}} t_{j_l}, j = 1, J^n$ the *generalized times* for n-th order attacks (the order matters in the n-th order *attack models* so that we must consider all n-tuples. See below)

- $m = 0, M - 1$ the possible values of the *submessages*

- m^i the *submessage* for encryption i (plaintext or ciphertext (not exclusive))

- $r^{i,l} = 0, R-1,\ l = 1, n$ the masks for encryption i for countermeasures against n-th order attacks. We have $\forall i = 1, N,\ \overset{n}{\underset{l=1}{\otimes}} r^{i,l} = 0$ ($n-1$ degrees of freedom). E.g. $\otimes = \oplus$ for Boolean Masking

- $S_{t_j}^i, j = 1, J$ the *side-channel signal* for encryption i at *time* t_j : discrete-time and discrete-value scalar or vector field (e.g. EMA attacks) stochastic processes. However, following a long tradition, we will nevertheless consider continuous stochastic processes (e.g. Gaussian processes) in a first step. But ultimately, we need to deal only with discrete direct probability distributions with small compact support

- $S_{t_j^n}^i = \overset{n}{\underset{l=1}{\wedge}} S_{t_{j_l}}^i, j = 1, J^n$ the *generalized side-channel signal* for encryption i at *generalized time* t_j^n for n-th order attacks

- $D_{t_j^n}^i$ the n-th order *attack datum* for encryption i at *generalized time* t_j^n. Examples:

    - first-order *"DPA-like" attack datum*: $D_{t_j}^i = m^i \wedge S_{t_j}^i$
    - n-th order *"DPA-like" attack datum*: $D_{t_j^n}^i = m^i \wedge S_{t_j^n}^i$
    - n-th-order *"Template-like" attack datum*: $D_{t_j^n}^i = k_0 \wedge m^i \wedge S_{t_j^n}^i$ (known fixed subkey), $D_{t_j^n}^i = k^i \wedge m_0 \wedge S_{t_j^n}^i$ (known/chosen subkeys, fixed submessage. E.g. key-scheduling attacks) or more generally $D_{t_j^n}^i = k^i \wedge m^i \wedge S_{t_j^n}^i$

In this paper, we give examples for n-th order *"DPA-like" attacks* with *datum* $D_{t_j^n}^i = m^i \wedge S_{t_j^n}^i$. But it shall be clear that the calculations are essentially the same for other kind of attacks.

- $D^A_{t^n_j} = \overset{N}{\underset{i=1}{\wedge}} D^i_{t^n_j}$ the n-th order *attack data* at *generalized time* t^n_j

- $D^A = \overset{N}{\underset{i=1}{\wedge}} \overset{J^n}{\underset{j=1}{\wedge}} D^i_{t^n_j}$ the n-th order *attack data* for attacker A

- $F_d(m^i, k^i), d = 1, D$ the *decision functions* (intermediate variables of the cipher implementation) on which the attack is mounted. Important special cases are *decision functions* $F_d(m^i, k)$ for one-shot *"DPA-like" attacks* and $F_d(k^i)$ or $F_d\left(m^i, k_0\right)$ or $F_d\left(m^i, k^i\right)$ for the *profiling phase* of some Template Attacks (e.g. $F_d(k^i) = k^i$ for some key-scheduling attacks)

- $\mathrm{M}^{iL}_{t^n_j} = {}'S^i_{t^n_j} = \Phi(W, \Theta^i_{t^n_j})'$ a n-th order parametric *leakage model* with *hyperparameters/nuisance parameters* $\Theta^i_{t^n_j}$, describing the Attacker *state of knowledge* on the causal link between the logical word W processed by the target device at *generalized time* t^n_j and the *generalized side-channel signal* $S^i_{t^n_j}$. If the hyperparameters $\Theta^i_{t^n_j}$ are the same for all i (e.g. first-order attacks), we note simply $\mathrm{M}^L_{t^n_j} = {}'S_{t^n_j} = \Phi(W, \Theta_{t^n_j})'$. Examples:

  - First-order scalar Hamming-Laplace-Gauss *leakage model* with algorithmic noise and unknown constant initial state W_0

$$\begin{cases} \mathrm{M}^{HL}_{t_j} = {}'S_{t_j} = \varepsilon_{t_j} h\left(W \oplus W_0\right) + \mu^1_{t_j} + \xi^1_{t_j} + \xi^2_{t_j} * B_{t_j}\,' \\ \xi^1_{t_j} \sim N\left[0, \left(\sigma^1_{t_j}\right)^2\right] \quad \xi^2_{t_j} \sim N\left[\mu^2_{t_j}, \left(\sigma^2_{t_j}\right)^2\right] \quad B_{t_j} \text{ pseudo-binomial distribution} \\ \Theta^H_{t_j} = \varepsilon_{t_j} \wedge \mu^1_{t_j} \wedge \sigma^1_{t_j} \wedge \mu^2_{t_j} \wedge \sigma^2_{t_j} \wedge W_0 \ldots \\ h(.) \text{ Hamming weight function} \end{cases}$$

  - n-th order scalar Hamming-Laplace-Gauss *leakage model* with masking

$$\mathrm{M}^{iL}_{t^n_j} = {}'S^i_{t_{j1}} = \varepsilon_{t_{j1}} h\left(W \otimes r^{i,1}\right) + \mu_{t_{j1}} + \xi^1_{t_{j1}} * B + \xi^2_{t_{j1}}\,' \wedge \ldots \wedge {}'S^i_{t_{jn}} = \varepsilon_{t_n} h\left(r^{i,n}\right) + \mu_{t_{jn}} + \xi^1_{t_{jn}} * B + \xi^2_{t_{jn}}\,'$$

  - First-order bitwise scalar Laplace-Gauss *leakage model* without algorithmic noise for a C-bits logical word $W = \sum_{c=0}^{C-1} b_c 2^c$

$$\begin{cases} \mathrm{M}_{t_j}^{L} = 'S_{t_j} = \sum_{c=1}^{C} \alpha_{t_j}^{c} b_c + \mu_{t_j} + \xi_{t_j}\, ' \\ \xi_{t_j} \sim N\left(0, \sigma_{t_j}{}^{2}\right) \\ \Theta_{t_j} = \mu_{t_j} \wedge \sigma_{t_j} \wedge \bigwedge_{c=1}^{C} \alpha_{t_j}^{c} \end{cases}$$

- $\mathrm{M}_{t_j^n}^{iA} = 'S_{t_j^n}^{i} = \Phi\left[F\left(m^i, k^i\right), \Theta_{t_j^n}^{i}\right]'$ the n -th order *attack model* for *decision function* $F(m^i, k^i)$, *attack datum* $D_{t_j^n}^{i} = k^i \wedge m^i \wedge S_{t_j^n}^{i}$ and *leakage model* $\mathrm{M}_{t_j^n}^{iL} = 'S_{t_j^n}^{i} = \Phi(W, \Theta_{t_j^n}^{i})'$.

- $\mathrm{M}_{t_j^n}^{i\neg A} = 'S_{t_j^n}^{i} = \Psi(W^i, \mathrm{N}_{t_j^n}^{i})'$ a n -th order parametric *"non-attack" model* describing the Attacker *state of knowledge* on the *generalized side-channel signal* $S_{t_j^n}^{i}$ at *generalized time* t_j^n if no *decision function* $F\left(m^i, k^i\right)$ of interest is processed. Examples:

    - $\Psi = \Phi$ (same underlying *leakage model*), $W = \text{constant}$: constant, noisy signal. E.g. Gaussian signal

    - $\Psi = \Phi$, $W^i = m^i$: message processing. E.g. Gaussian signal conditional upon m .

    - $\Psi = \Phi$, $W^i = \text{unknown logical word}$: algorithmic noise. E.g.: signal whose probability density function is equal to the convolution product of a Laplace-Gauss distribution and a pseudo-binomial one (for the Hamming weights).

II Preparing the General Solution: Solving Two Sub-Problems

Let us revisit first, from our new point of view, standard, generic known-messages side-channel attacks on symmetric block ciphers such as DPA [10][32][47][64][72][81], CPA [10][21][23][32][67][72][81], MIA [5][30][81], Stochastic Methods [31][76], Template Attacks [1][2][22][31][32][80], PCA [2][80], LDA [80], high-order attacks [10][32][35][44][57][65], etc.

Clearly, we can identify two ubiquitous, general and generic problems:

- **Problem 1: find out the *generalized leakage times* at which one or several *decision functions* of interest is/are processed**

- **Problem 2: find out the subkey, given *generalized leakage times***

For example, **Problem 1** corresponds typically to the *profiling phase* of Template Attacks while **Problem 2** corresponds to their *extraction phase* [1][2][31][32][80].

So, let us first provide formal, general solutions to both problems independently.

1) Problem 1

First, let us solve this problem for a single *decision function* $F\left(m^i,k^i\right)$. In section III, we will generalize this solution to an arbitrary number of *decision functions.*

Problem 1 is clearly a set of Hypotheses Testing problems indexed by the *generalized time*. At each *generalized time* t_j^n, we want to test the null hypothesis

$$H_{t_j^n}^0 : \text{ there exists side-channel leakage on decision function } F\left(m^i,k^i\right)$$

against the alternative

$$H_{t_j^n}^1 = \neg H_{t_j^n}^0 : \text{ there exists no side-channel leakage on decision function } F\left(m^i,k^i\right)$$

Generally speaking, this kind of problem can be tackled in two main statistical frameworks: the parametric and the non-parametric ones [74]. In the parametric framework, we are provided with a parametric model under both hypotheses and in the nonparametric one, we are not. As we said before, since *Probability as Logic* describes only *states of knowledge*, the parametric setting does not mean at all that we **assume** our data to follow those parametric models/distributions, but only that the *corpus of knowledge/background information* sufficiently constraints probability distributions to fall into finite-dimensional manifolds.

Parametric Solution

By contrast to some non-Bayesian, orthodox hypotheses tests (e.g. *omnibus* tests, Fisher P-values.) [40][62][63], we must always specify explicitly the alternative hypothesis $H_{t_j^n}^1$ within

Probability as Logic. Otherwise, by the *Theorem of Total Probability*, the null hypothesis $H^0_{t^n_j}$ has probability 1 and for sure there is no problem at all.

So, in addition to side-channel *attack models* for the null hypothesis $H^0_{t^n_j}$, we must also provide side-channel *"non-attack" models* for the alternative hypothesis $H^1_{t^n_j}$, describing our state of knowledge on the side-channel signals if the *decision function* $F\left(m^i, k^i\right)$ of interest is not processed. This is a first new point compared to all procedures proposed so far.

We say model**s** and not model because another key point for understanding Side-channel Cryptanalysis is to recognize that we have, generally speaking, one model for each particular side-channel *attack datum* $D^i_{t^n_j}$ at each *generalized time* t^n_j, not only a single model for the whole *attack data* $D^A_{t^n_j} = \overset{N}{\underset{i=1}{\wedge}} D^i_{t^n_j}$ at each *generalized time*. This is typically the case for n-th order attacks [10][32][35][44][57][65]: for each side-channel signal we have $n-1$ new masks. They are unknown hyperparameters entering in our problem that we must estimate together with the subkey and the physical hyperparameters. So, we must plug them in our models. As a result, we do have a new model for each n-th order side-channel *attack datum* $D^i_{t^n_j}, i = 1, N$.

So, let $\mathrm{M}^{iA}_{t^n_j}$ and $\mathrm{M}^{i\neg A}_{t^n_j}$ be the *attack datum model* and the *"non-attack" datum model* respectively for *attack datum* $D^i_{t^n_j}$ with hyperparameters $\Theta^i_{t^n_j}$ and $\mathrm{N}^i_{t^n_j}$. Then, we have the *attack data model* $\mathrm{M}^A_{t^n_j} = \overset{N}{\underset{i=1}{\wedge}} \mathrm{M}^{iA}_{t^n_j}$ and the *"non-attack" data model* $\mathrm{M}^{\neg A}_{t^n_j} = \overset{N}{\underset{i=1}{\wedge}} \mathrm{M}^{i\neg A}_{t^n_j}$ for *attack data* $D^A_{t^n_j} = \overset{N}{\underset{i=1}{\wedge}} D^i_{t^n_j}$ with hyperparameters $\Theta_{t^n_j} = \overset{N}{\underset{i=1}{\wedge}} \Theta^i_{t^n_j}$ and $\mathrm{N}_{t^n_j} = \overset{N}{\underset{i=1}{\wedge}} \mathrm{N}^i_{t^n_j}$ respectively.

Then, **Problem 1** admits a formal solution, as a theorem of *Probability as Logic*. What we really want are the posterior probabilities for our hypotheses

$$p\left(H^0_{t^n_j} \middle| D^A_{t^n_j} \wedge I^A\right) = p\left(\mathrm{M}^A_{t^n_j} \middle| D^i_{t^n_j} \wedge I^A\right) \text{ and } p\left(H^1_{t^n_j} \middle| D^A_{t^n_j} \wedge I^A\right) = p\left(\mathrm{M}^{\neg A}_{t^n_j} \middle| D^A_{t^n_j} \wedge I^A\right) = 1 - p\left(H^0_{t^n_j} \middle| D^A_{t^n_j} \wedge I^A\right)$$

that are given respectively by Bayes Rule/Theorem

$$p\left(\mathrm{M}^A_{t^n_j} \middle| D^A_{t^n_j} \wedge I^A\right) = \frac{p\left(\mathrm{M}^A_{t^n_j} \middle| I^A\right) p\left(D^A_{t^n_j} \middle| \mathrm{M}^A_{t^n_j} \wedge I^A\right)}{p\left(D^A_{t^n_j} \middle| \mathrm{M}^A_{t^n_j} \wedge I^A\right) p\left(\mathrm{M}^A_{t^n_j} \middle| I^A\right) + p\left(D^A_{t^n_j} \middle| \mathrm{M}^{\neg A}_{t^n_j} \wedge I^A\right) p\left(\mathrm{M}^{\neg A}_{t^n_j} \middle| I^A\right)}$$

and

$$p\left(\mathrm{M}^{\neg A}_{t^n_j} \middle| D^A_{t^n_j} \wedge I^A\right) = \frac{p\left(\mathrm{M}^{\neg A}_{t^n_j} \middle| I^A\right) p\left(D^A_{t^n_j} \middle| \mathrm{M}^{\neg A}_{t^n_j} \wedge I^A\right)}{p\left(D^A_{t^n_j} \middle| \mathrm{M}^A_{t^n_j} \wedge I^A\right) p\left(\mathrm{M}^A_{t^n_j} \middle| I^A\right) + p\left(D^A_{t^n_j} \middle| \mathrm{M}^{\neg A}_{t^n_j} \wedge I^A\right) p\left(\mathrm{M}^{\neg A}_{t^n_j} \middle| I^A\right)}$$

However, it is common practice to report rather the so-called *Bayes Factors* [7][40][74][83]

$$B_{t_j^n}^{A\neg A}=\frac{p\left(D_{t_j^n}^A\middle|\mathrm{M}_{t_j^n}^A\wedge I^A\right)}{p\left(D_{t_j^n}^A\middle|\mathrm{M}_{t_j^n}^{\neg A}\wedge I^A\right)}=\frac{p\left(\mathrm{M}_{t_j^n}^A\middle|D_{t_j^n}^A\wedge I^A\right)}{p\left(\mathrm{M}_{t_j^n}^{\neg A}\middle|D_{t_j^n}^A\wedge I^A\right)}/\frac{p\left(\mathrm{M}_{t_j^n}^A\middle|I^A\right)}{p\left(\mathrm{M}_{t_j^n}^{\neg A}\middle|I^A\right)}$$

because we do not need to supply the prior probabilities of the models $p\left(\mathrm{M}_{t_j^n}^A\middle|I^A\right)$ and $p\left(\mathrm{M}_{t_j^n}^{\neg A}\middle|I^A\right)$. As a consequence, those Bayes Factors just quantify in what extent the data $D_{t_j^n}^A$ support the null model against the alternative and provide us with a more "objective" procedure if there is ever some arbitrariness, some indeterminacy in those prior probabilities.

By the *Theorem of Total Probability*, we have

$$B_{t_j^n}^{A\neg A}=\frac{p\left(D_{t_j^n}^A\middle|\mathrm{M}_{t_j^n}^A\wedge I^A\right)}{p\left(D_{t_j^n}^A\middle|\mathrm{M}_{t_j^n}^{\neg A}\wedge I^A\right)}=\frac{\sum_{k=0}^{K-1}\int_{\Theta_{t_j^n}}p\left(D_{t_j^n}^A\middle|k\wedge\Theta_{t_j^n}\wedge\mathrm{M}_{t_j^n}^A\wedge I^A\right)p\left(k\wedge\Theta_{t_j^n}\middle|\mathrm{M}_{t_j^n}^A\wedge I^A\right)d\Theta_{t_j^n}}{\int_{\mathrm{N}_{t_j^n}}p\left(D_{t_j^n}^A\middle|\mathrm{N}_{t_j^n}\wedge\mathrm{M}_{t_j^n}^{\neg A}\wedge I^A\right)p\left(\mathrm{N}_{t_j^n}\middle|\mathrm{M}_{t_j^n}^{\neg A}\wedge I^A\right)d\mathrm{N}_{t_j^n}}$$

An important particular case is when subkey k is known *a priori* to be equal to k_0 $\Leftrightarrow p\left(k\middle|\mathrm{M}_{t_j^n}^A\wedge I^A\right)=\delta\left(k-k_0\right)$, for instance in the *profiling phase* of some *Template Attacks*

$$B_{t_j^n}^{A\neg A}=\frac{p\left(D_{t_j^n}^A\middle|\mathrm{M}_{t_j^n}^A\wedge I^A\right)}{p\left(D_{t_j^n}^A\middle|\mathrm{M}_{t_j^n}^{\neg A}\wedge I^A\right)}=\frac{\int_{\Theta_{t_j^n}}p\left(D_{t_j^n}^A\middle|k_0\wedge\Theta_{t_j^n}\wedge\mathrm{M}_{t_j^n}^A\wedge I^A\right)p\left(\Theta_{t_j^n}\middle|\mathrm{M}_{t_j^n}^A\wedge I^A\right)d\Theta_{t_j^n}}{\int_{\mathrm{N}_{t_j^n}}p\left(D_{t_j^n}^A\middle|\mathrm{N}_{t_j^n}\wedge\mathrm{M}_{t_j^n}^{\neg A}\wedge I^A\right)p\left(\mathrm{N}_{t_j^n}\middle|\mathrm{M}_{t_j^n}^{\neg A}\wedge I^A\right)d\mathrm{N}_{t_j^n}}$$

Examples:

- **First-order Hamming-Laplace-Gauss "DPA-like" attack with constant, noisy Gaussian "non-attack" model**.

Consider the basic first-order Hamming-Laplace-Gauss side-channel *attack data model*

$$\mathrm{M}_{t_j}^{AH}='S_{t_j}^i=\varepsilon_{t_j}h\left[F\left(m^i,k\right)\right]+\mu_{t_j}+\xi_{t_j}'$$

with $\xi_{t_j}\sim N\left(0,\sigma_{t_j}{}^2\right)$. The hyperparameters are

$$\Theta_{t_j}^H=\varepsilon_{t_j}\wedge\mu_{t_j}\wedge\sigma_{t_j}$$

that are the same for all *attack datum*. Consider the first-order *"DPA-like" attack data*

$$D_{t_j}^{A} = \overset{N}{\underset{i=1}{\wedge}} D_{t_j}^{i} = \overset{N}{\underset{i=1}{\wedge}} m^{i} \wedge S_{t_j}^{i}$$

For known-messages attacks with exchangeable data, the direct probabilities of the attack data given the parameters and the model, or likelihood, write as

$$\begin{aligned} p\left(D_{t_j}^{A} \middle| k \wedge \Theta_{t_j} \wedge \mathrm{M}_{t_j}^{A} \wedge I^{A}\right) &= p\left(\overset{N}{\underset{i=1}{\wedge}} m^{i} \wedge S_{t_j}^{i} \middle| k \wedge \Theta_{t_j} \wedge \mathrm{M}_{t_j}^{A} \wedge I^{A}\right) \\ &= \prod_{i=1}^{N} p\left(m^{i} \wedge S_{t_j}^{i} \middle| k \wedge \Theta_{t_j} \wedge \mathrm{M}_{t_j}^{A} \wedge I^{A}\right) \\ &= \prod_{i=1}^{N} p\left(S_{t_j}^{i} \middle| m^{i} \wedge k \wedge \Theta_{t_j} \wedge \mathrm{M}_{t_j}^{A} \wedge I^{A}\right) p\left(m^{i} \middle| k \wedge \Theta_{t_j} \wedge \mathrm{M}_{t_j}^{A} \wedge I^{A}\right) \end{aligned}$$

Thus, for model $\mathrm{M}_{t_j}^{AH}$ and uniformly distributed known, non-chosen *submessages* we have

$$p\left(D_{t}^{A} \middle| k \wedge \Theta_{t_j}^{H} \wedge M_{t_j}^{AH} \wedge I^{A}\right) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_{t_t}} e^{-\frac{\left(S_{t_j}^{i} - \varepsilon_{t_j} h_{k}^{i} - \mu_{t_j}\right)^{2}}{2\sigma_{t_j}{}^{2}}} M^{-1} = \left(2\pi\right)^{-\frac{N}{2}} M^{-N} \sigma_{t_j}{}^{-N} e^{-\frac{1}{2\sigma_{t_j}{}^{2}} \sum_{i=1}^{N} \left(S_{t_j}^{i} - \varepsilon_{t_j} h_{k}^{i} - \mu_{t_j}\right)^{2}}$$

where

$$h_{k}^{i} = h\left[F\left(m^{i}, k\right)\right]$$

Consider also the Gaussian *"non-attack" data model* for constant but noisy side-channel signals

$$\mathrm{M}_{t_j}^{\neg AG} = {}'S_{t_j}^{i} = \mu_{t_j}{}' + \xi_{t_j}{}''$$

with $\xi_{t_j}{}' \sim N\left(0, \sigma_{t_j}{}'^{2}\right)$ and hyperparameters

$$\mathrm{N}_{t}^{G} = \mu_{t}{}' \wedge \sigma_{t}{}'$$

The direct probabilities of the *side-channel data* given the parameters and the *"non-attack" model* write as

$$p\left(D_{t_j}^{A} \middle| \mathrm{N}_{t_j} \wedge \mathrm{M}_{t_j}^{\neg AG} \wedge I^{A}\right) = \prod_{i=1}^{N} p\left(S_{t_j}^{i} \middle| \mathrm{N}_{t_j} \wedge \mathrm{M}_{t_j}^{\neg AG} \wedge I^{A}\right) p\left(m^{i} \middle| \mathrm{N}_{t_j} \wedge \mathrm{M}_{t_j}^{\neg AG} \wedge I^{A}\right)$$

For model $\mathrm{M}_{t_j}^{\neg AG}$ and uniformly distributed *submessages* we have

$$p\left(D_{t_k}^{A}\middle|\mathrm{N}_{t_j}^{G}\wedge\mathrm{M}_{t_j}^{\neg AG}\wedge I^{A}\right)=\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi}\sigma_{t_j}{}'}e^{-\frac{\left(S_{t_j}^{i}-\mu_{t_j}{}'\right)^{2}}{2\sigma_{t_j}{}'^{2}}}M^{-1}=(2\pi)^{-\frac{N}{2}}M^{-N}\sigma_{t_j}{}'^{-N}e^{-\frac{1}{2\sigma_{t_j}{}^{2}}\sum_{i=1}^{N}\left(S_{t_j}^{i}-\mu_{t_j}{}'\right)^{2}}$$

Therefore the Bayes factor writes as:

$$B_{t_j}^{A\neg A}=\frac{\sum_{k=0}^{K-1}\int_{\Theta_{t^n}}p\left(D_{t_j}^{A}\middle|k\wedge\Theta_{t_j}^{H}\wedge\mathrm{M}_{t_j}^{AH}\wedge I^{A}\right)p\left(k\wedge\Theta_{t_j}^{H}\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{A}\right)d\Theta_{t^n}}{\int_{\mathrm{N}_{t^n}}p\left(D_{t_j}^{A}\middle|\mathrm{N}_{t}^{G}\wedge\mathrm{M}_{t_j}^{\neg AG}\wedge I^{A}\right)p\left(\mathrm{N}_{t_j}^{G}\middle|\mathrm{M}_{t_j}^{\neg AG}\wedge I^{A}\right)d\mathrm{N}_{t^n}}=$$

$$\frac{\sum_{k=0}^{K-1}\int_{\sigma_{t_j}}\int_{\mu_{t_j}}\int_{\varepsilon_{t_j}}\sigma_{t_j}{}^{-N}e^{-\frac{1}{2\sigma_{t_j}{}^{2}}\sum_{i=1}^{N}\left(S_{t_j}^{i}-\varepsilon_{t_j}h_k^i-\mu_{t_j}\right)^{2}}p\left(k\wedge\varepsilon_{t_j}\wedge\mu_{t_j}\wedge\sigma_{t_j}\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{A}\right)d\varepsilon_{t_j}d\mu_{t_j}d\sigma_{t_j}}{\int_{\sigma_{t_j}{}'}\int_{\mu_{t_j}{}'}\sigma_{t_j}{}'^{-N}e^{-\frac{1}{2\sigma_{t_j}{}'^{2}}\sum_{i=1}^{N}\left(S_{t_j}^{i}-\mu_{t_j}{}'\right)^{2}}p\left(\mu_{t_j}{}'\wedge\sigma_{t_j}{}'\middle|\mathrm{M}_{t_j}^{\neg AG}\wedge I^{A}\right)d\mu_{t_j}{}'d\sigma_{t_j}{}'}$$

We do not deal with the assignment of joint prior probability distributions for the parameters such as

$$p\left(k\wedge\varepsilon_{t_j}\wedge\mu_{t_j}\wedge\sigma_{t_j}\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{A}\right)\qquad\text{and}\qquad p\left(\mu_{t_j}{}'\wedge\sigma_{t_j}{}'\middle|\mathrm{M}_{t_j}^{\neg AG}\wedge I^{A}\right)$$

in this paper. Let us nevertheless finish the calculation only once for simple joint prior distributions to see what kind of solution we can expect at the end in general.

Consider the silly but fairly non-informative Bayes-Laplace-Lhoste-Jeffreys improper prior distribution for an ignorant Attacker *A* with background information I^{IA} (e.g. he has never heard about ISO 7816, he does not even know that the static power consumption μ_{t_j} is positive, etc.)

$$\begin{aligned}p\left(k\wedge\varepsilon_{t_j}\wedge\mu_{t_j}\wedge\sigma_{t_j}\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{IA}\right)&=p\left(\varepsilon_{t_j}\wedge\mu_{t_j}\wedge\sigma_{t_j}\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{IA}\right)p\left(k\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{IA}\right)\\&=p\left(\mu_{t_j}\wedge\sigma_{t_j}\middle|\varepsilon_{t_j}\wedge\mathrm{M}_{t_j}^{AH}\wedge I^{IA}\right)p\left(\varepsilon_{t_j}\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{IA}\right)p\left(k\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{IA}\right)\\&=p\left(\mu_{t_j}\wedge\sigma_{t_j}\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{IA}\right)p\left(\varepsilon_{t_j}\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{IA}\right)p\left(k\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{IA}\right)\\&\propto K^{-1}\sigma_{t_j}{}^{-1}\end{aligned}$$

on $\{0,...,K-1\}\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}^{+*}$ and the translation- and scale-invariant Lhoste-Jeffreys prior for the Laplace-Gauss distribution [7][40][43][55][74]

$$p\left(\mu_{t_j}{}'\wedge\sigma_{t_j}{}'\middle|\mathrm{M}_{t_j}^{\neg AG}\wedge I^{IA}\right)\propto\sigma_{t_j}{}'^{-1}$$

on $\mathbb{R}\times\mathbb{R}^{+*}$. ($p\left(\varepsilon_{t_j}\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{AI}\right)$ should be a Gaussian and not a flat Bayes-Laplace prior for a fully invariant prior [58].).

Using the identity

$$\int_{-\infty}^{+\infty} e^{-a\theta^2+b\theta}d\theta=\pi^{\frac{1}{2}}a^{-\frac{1}{2}}e^{\frac{b^2}{4a}} \qquad\qquad a>0$$

for computing both ε_{t_j} and μ_{t_j} Gaussian integrals and the change of variable

$$\sigma_{t_j}=Q_{t_j}^{k\,\frac{1}{2}}x_{t_j}^{-\frac{1}{2}} \qquad x_{t_j}=Q_{t_j}^{k}\sigma_{t_j}^{-2} \qquad d\sigma_{t_j}=-\frac{1}{2}Q_{t_j}^{k\,\frac{1}{2}}x_{t_j}^{-\frac{3}{2}}dx_{t_j} \qquad Q_{t_j}^{k}=\frac{N}{2}\hat{\sigma}_{S_{t_j}}^{\ 2}\left(1-\hat{\rho}_{h_k,S_{t_j}}^{\ 2}\right)$$

for computing the σ_{t_j} Eulerian $\Gamma-$ integral of the second kind, we find, after some algebra, the marginal probability distribution for *attack data* $D_{t_j}^{A}$ conditional on the *attack model* $M_{t_j}^{AH}$

$$\forall N>2,\quad p\left(D_{t_j}^{A}\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{IA}\right)=$$

$$(2\pi)^{-\frac{N}{2}}M^{-N}K^{-1}\sum_{k=0}^{K-1}\int_{0}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\sigma_{t_j}^{\ -N-1}e^{-\frac{1}{2\sigma_{t_j}^{\ 2}}\sum_{i=1}^{N}\left(S_{t_j}^{i}-\varepsilon_{t_j}h_k^i-\mu_{t_j}\right)^2}d\varepsilon_{t_j}d\mu_{t_j}d\sigma_{t_j}=$$

$$\frac{1}{2}(\pi)^{\frac{2-N}{2}}\Gamma\left(\frac{N-2}{2}\right)N^{-\frac{N}{2}}M^{-N}K^{-1}\hat{\sigma}_{S_{t_j}}^{\ 2-N}\sum_{k=0}^{\mathrm{K}-1}\hat{\sigma}_{h_k}^{\ -1}\left(1-\hat{\rho}_{h_k,S_{t_j}}^{\ 2}\right)^{\frac{2-N}{2}}\propto$$

$$\hat{\sigma}_{S_{t_j}}^{\ 2-N}\sum_{k=0}^{\mathrm{K}-1}\hat{\sigma}_{h_k}^{\ -1}\left(1-\hat{\rho}_{h_k,S_{t_j}}^{\ 2}\right)^{\frac{2-N}{2}}$$

where $\hat{\sigma}_{S_{t_j}}$ stands for the *empirical/sample standard deviation* of the side-channel signals at time t_j, $\hat{\sigma}_{h_k}$ for the sample standard deviation of the Hamming weights of the *decision function* for subkey k and $\hat{\rho}_{h_k,S_{t_j}}^{\ 2}$ for K. Pearson *sample determination coefficient* between those Hamming weights and the side-channel signals (the *determination coefficient* is just the square of the *correlation coefficient*. See below.).

We can recognize a mixture of Student *t* distributions over subkeys *k* [17]. For $N\leq 2$, the integrals are divergent: this is how *Probability as Logic* warns us that we do not have enough information to answer the question we asked her [40].

In the same way, we find the marginal probability distribution for data $D_{t_j}^{A}$ conditional on the *"non-attack" model* $\mathrm{M}_{t_j}^{\neg AG}$:

$$p\left(D_{t_j}^{A}\middle|\wedge \mathrm{M}_{t_j}^{\neg AG}\wedge I^{IA}\right)=$$

$$(2\pi)^{-\frac{N}{2}}M^{-N}\int_{0}^{+\infty}\int_{-\infty}^{+\infty}\sigma_{t_j}{'}^{-N-1}e^{-\frac{1}{2\sigma_{t_j}{'}^{2}}\sum_{i=1}^{N}\left(S_{t_j}^{i}-\mu_{t_j}{'}\right)^{2}}d\mu_{t_j}{'}d\sigma_{t_j}{'}\propto$$

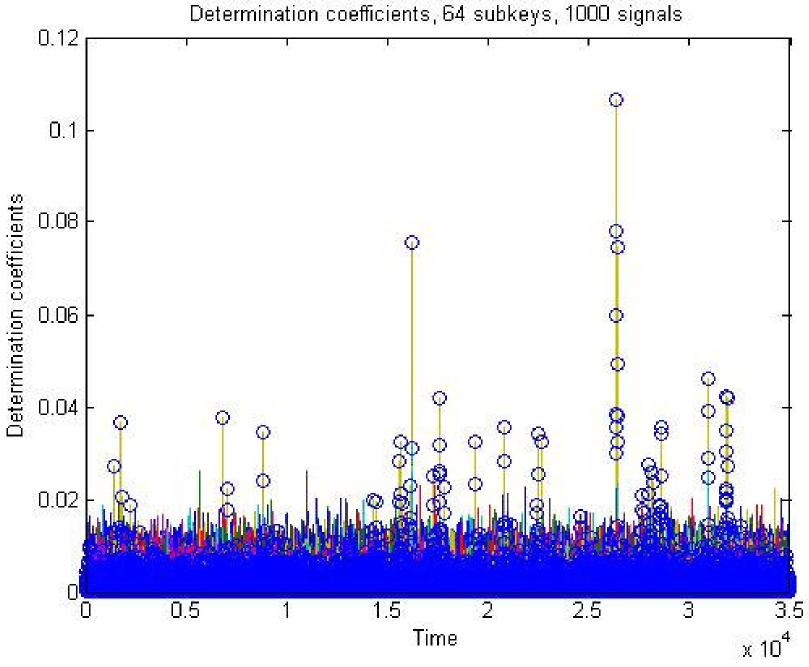
$$\hat{\sigma}_{S_{t_j}}{}^{1-N}$$

Finally, the Bayes Factors for this ignorant Attacker and those Gaussian models writes as

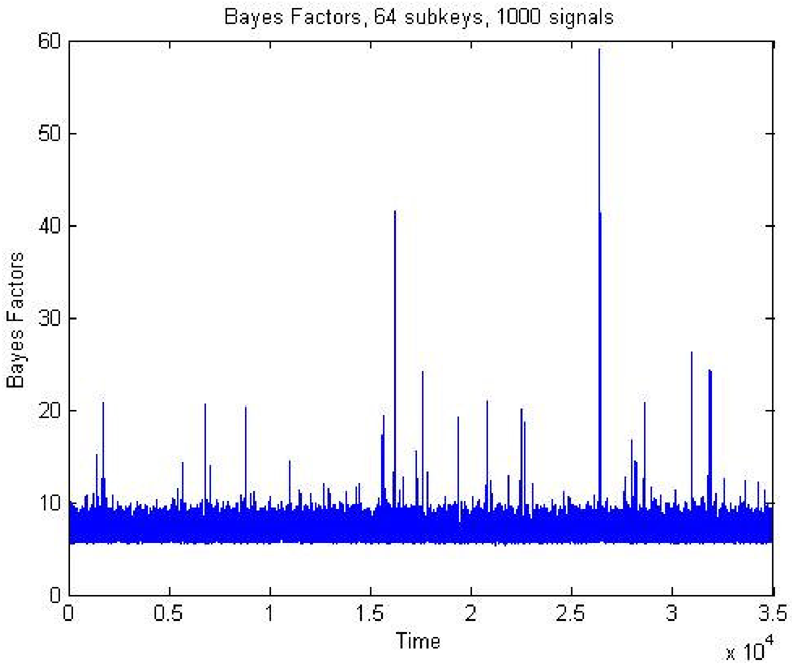
$$\forall N>2,\quad B_{t_j}^{HG}=\frac{p\left(D_{t_j}^{A}\middle|\mathrm{M}_{t_j}^{AH}\wedge I^{IA}\right)}{p\left(D_{t_j}^{A}\middle|\mathrm{M}_{t_j}^{\neg AG}\wedge I^{IA}\right)}\propto\hat{\sigma}_{S_{t_j}}\sum_{k=0}^{\mathrm{K}-1}\hat{\sigma}_{h_k}{}^{-1}\left(1-\hat{\rho}_{h_k,S_{t_j}}{}^{2}\right)^{\frac{2-N}{2}}$$

Here are some experimental results for an unprotected software DES implementation on an old 8 bits microcontroller. The *"DPA-like" decision function* $F\left(m^{i},k\right)$ is the 4 output bits of one DES S-box.

With 1000 total power consumption signals, a first-order CPA attack [10][21][23][32][67][72][81], based on the *determination coefficients* for all 2^6 possible values of the subkey, gives



The attack is clearly successful: we have significant correlation/determination peaks for one and only one particular value of the subkey (in yellow), which is the true value k_0 (blue circles). With the same data, the Bayes Factors procedure gives, in logarithmic scale



The Bayes Factors peaks are similar to determination coefficients ones. But at the same time, the amplitude of the *background correlation noise*, when the *decision function* is not processed, is clearly reduced. A possible measure of the discrimination power or stringency of those procedures is the Signal-to-Noise-Ratio between the maximum amplitude of the "correlation peaks" and the average amplitude of the background correlation noise. For CPA, we have roughly

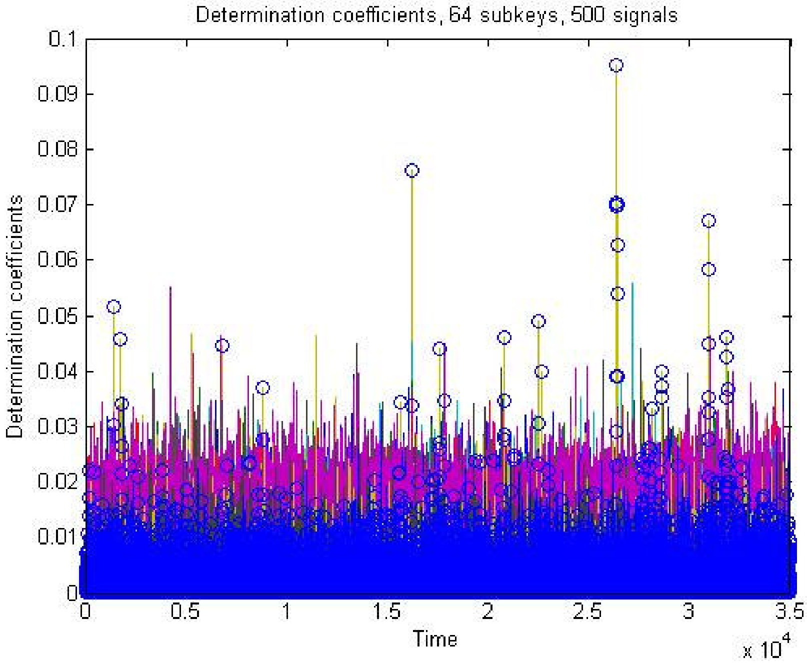
$$SNR_{CPA} = \frac{0.11}{0.017} = 6.5$$

and for the Bayes Factors procedure

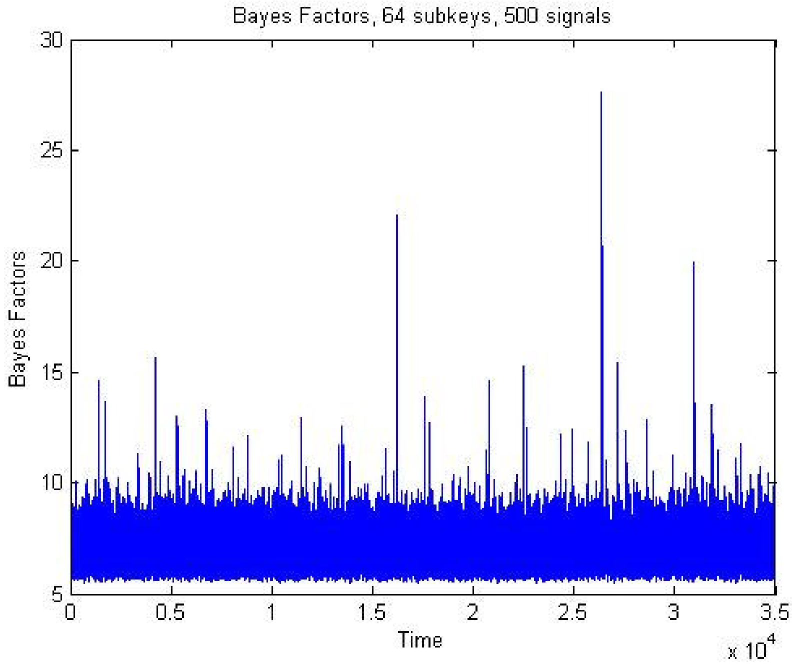
$$SNR_{BF} = \frac{59-6}{4} = 13$$

The SNR for the Bayes Factors Procedure is just twice the SNR for CPA.

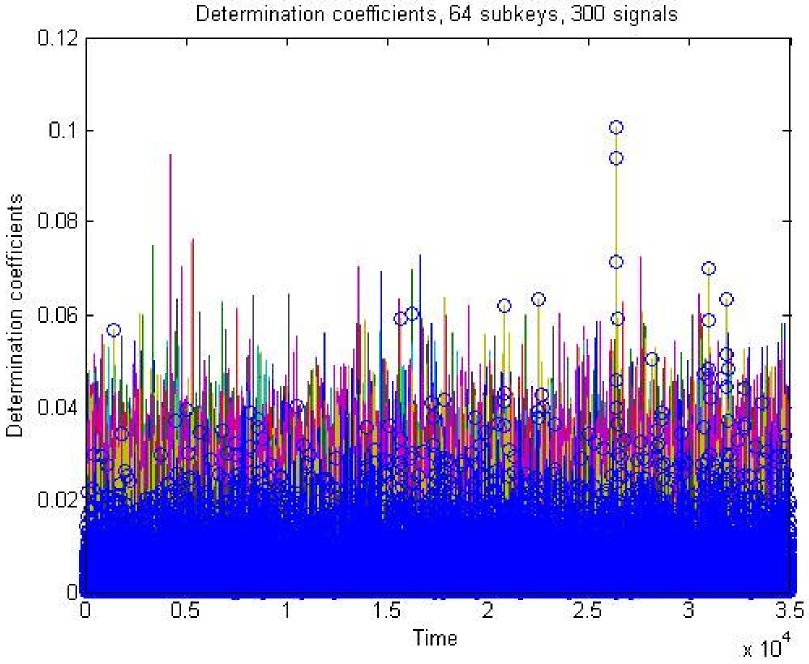
With 500 signals, CPA gives



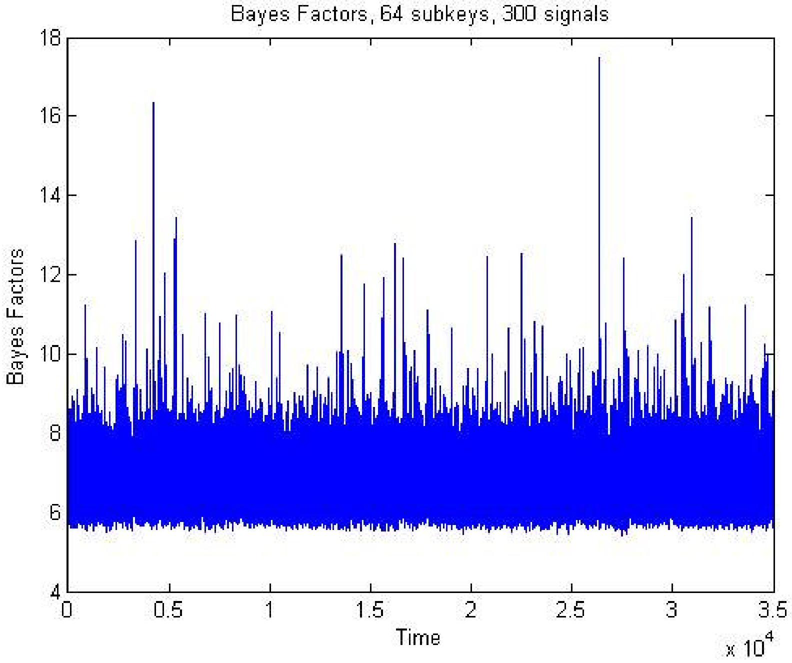
and the Bayes Factors procedure gives



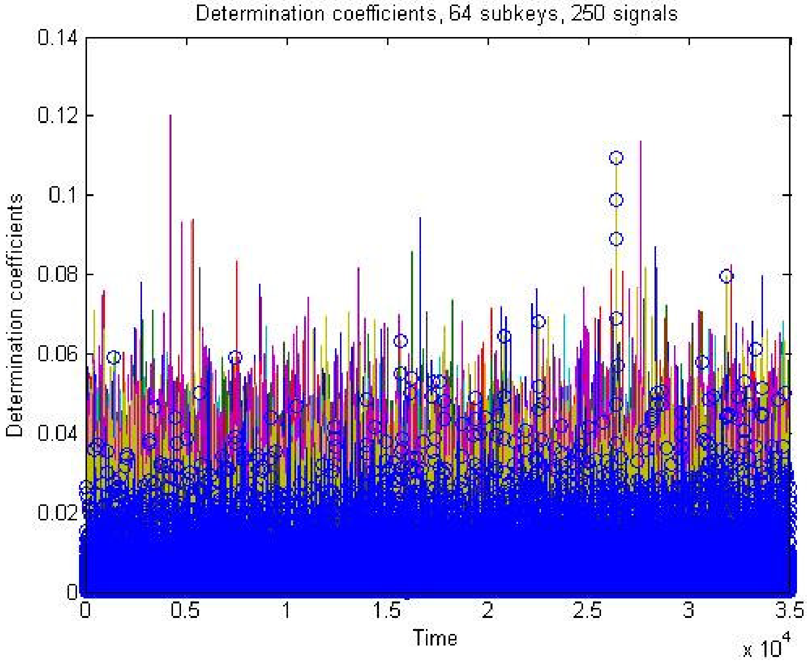
The implementation is still weakly broken by CPA with only 300 signals



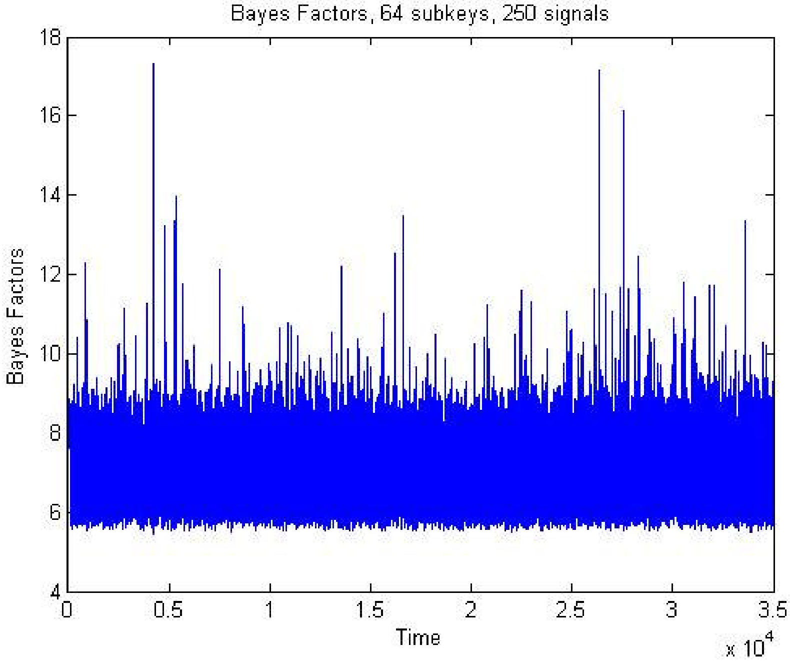
and the Bayes Factors still show us the most leaking *time*:



With 250 signals, the CPA attack is no more successful



as well as the Bayes Factors procedure



Those preliminary results are encouraging and validate, in some extent, our new approach: if the "correlation peaks" are asymptotically roughly the same in both procedures, the background correlation noise is significantly reduced in the Bayes Factors one. However, in the present case, it seems that the Bayes Factors procedure does not allow the attacker to reduce significantly the number of signals he needs in order to determine the *leakage times* with significant probabilities.

From now on the main point is that we have a general and automatic procedure applicable to any *attack datum* and *data*, any *attack model*, any *"non-attack" model* and any joint prior probability distributions for the parameters of both models. Clearly, the Gaussian *"non-attack" model* and the joint prior probability distributions for the parameters that we used in our experiments are the most non-informative, the least accurate and silliest we can imagine. We used them only for analytic convenience. So, it must be clear that those preliminary experimental results are in fact the **worst** ones we can get with this new Bayes Factors procedure. And they are already not so bad, at least not worst than CPA *for the same underlying Hamming-Laplace-Gauss attack model* (see below for details.).

Even in this simple, poor setting, we would need dedicated numerical algorithms in order to compute quickly and accurately the sums of the very high powers in the Bayes Factors. Clearly, since side-channel attack data are integer-valued and direct probability distributions are discrete in practice, we should try to work as much as possible with integer fractions in order to compute those very high powers exactly and to avoid round-off errors. A full symbolic and numerical theory of Side-channel Cryptanalysis is therefore waiting to be developed. Because we do not know how much our naïve Bayes Factors calculations suffer from such numerical errors, we unfortunately stop our preliminary experimental investigations at this early point.

Yet, let us sketch how we could compare and benchmark the Bayes Factors procedure to previous ones, from a theoretical point of view. This is not immediate and we find very interesting points on the way.

Under the Hamming-Laplace-Gauss *attack model*, Pearson linear correlation coefficient between the Hamming weights of the *decision function* and the signals within CPA attacks is supposed to be a relevant statistic for inferring *both leakage times and the subkey*. But the link between this particular side-channel *attack model* and this particular statistics is in fact not direct. To make it explicit, let us first derive Fisher *Profile Maximum Likelihood Estimator* (PMLE) [15] for subkey k that would be the standard non-Bayesian parametric statistic if *there would be one and only one leakage time*. This is the reason why time t is removed below.

As before, the likelihood writes as, now in Fisherian *ad hoc* non-probabilistic notations

$$L\left[k,\varepsilon,\mu,\sigma;(m,S)\right]=\prod_{i=1}^{N}P_M\left(m^i\right)\left(2\pi\sigma^2\right)^{-\frac{1}{2}}e^{-\frac{1}{2\sigma^2}\left(S^i-\varepsilon h_k^i-\mu\right)^2}$$

and the log-likelihood as

$$l_k=\ln L\left[k,\varepsilon,\mu,\sigma;(m,S)\right]=N\ln\left[\left(2\pi\sigma^2\right)^{-\frac{1}{2}}\right]-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(S^i-\varepsilon h_k^i-\mu\right)^2+\sum_{i=1}^{N}\ln P_M\left(m^i\right)$$

Cancelling first l_k partial derivates in ε, μ and σ, we find classical linear regression results

$$\hat{\varepsilon}_k=\frac{\hat{\sigma}_{h_k,S}}{\hat{\sigma}_{h_k}^2}\qquad\hat{\mu}_k=\overline{S}-\hat{\varepsilon}_k\overline{h_k}\qquad\text{and}\qquad\hat{\sigma}_k=\left[\frac{1}{N}\sum_{i=1}^{N}\left(S^i-\hat{\varepsilon}_k h_k^i-\hat{\mu}_k\right)^2\right]^{\frac{1}{2}}$$

where $\overline{S}$ stands for the *sample mean*, $\hat{\sigma}_{h_k}^2$ for the *sample variance*, $\hat{\sigma}_{h_k,S}$ for the *sample covariance* and $\hat{\varepsilon}_k$ for the *linear regression coefficient*. Plugging those estimates back into the likelihood, we get the *profile likelihoods* [15]

$$L\left[k,\hat{\varepsilon},\hat{\mu},\hat{\sigma};(m,S)\right] \propto \hat{\sigma}_k{}^{-N} \propto \left[\sum_{i=1}^{N}\left(S^i - \hat{\varepsilon}_k h_k^i - \hat{\mu}_k\right)^2\right]^{-\frac{N}{2}}$$

Using the identity for the sum of *least squares*

$$\sum_{i=1}^{N}\left(S^i - \hat{\varepsilon}_k h_k^i - \hat{\mu}_k\right)^2 = \hat{\sigma}_S{}^2\left(1-\hat{\rho}_{h_k,S}{}^2\right)$$

where $\hat{\rho}_{h_k,S} = \hat{\sigma}_{h_k,S}\hat{\sigma}_{h_k}{}^{-1}\hat{\sigma}_S{}^{-1}$ is K. Pearson *sample Product Moment Linear Correlation Coefficient,* Fisher *Profile Maximum Likelihood Estimator* of subkey k writes as

$$\hat{k} = \arg\max_k \left(1-\hat{\rho}_{h_k,S}{}^2\right)^{-\frac{N}{2}}$$

For a mono-bit *decision function* $B = F\left(M,k\right)$ (i.e. Kocher original *differential* attacks [47] or *partition attacks* [81]), the *determination coefficient* $\hat{\rho}_{b_k,S}{}^2$ reduces to the *point-biserial determination coefficient* [84]

$$\hat{\rho}_{b_k,S}{}^2 = \hat{\sigma}_{b_k,S}{}^2\hat{\sigma}_{b_k}{}^{-2}\hat{\sigma}_S{}^{-2} = \left(\overline{S|'b_k=1'} - \overline{S|'b_k=0'}\right)^2 \hat{\sigma}_S{}^{-2}\hat{\sigma}_{b_k}{}^2 \equiv K_k{}^2\hat{\sigma}_S{}^{-2}\hat{\sigma}_{b_k}{}^2$$

where K_k is (almost) the original differential DPA statistic [47] so that the *Profile Maximum Likelihood Estimator* is

$$\hat{k} = \arg\max_k \left(1-K_k{}^2\hat{\sigma}_S{}^{-2}\hat{\sigma}_{b_k}{}^2\right)^{-\frac{N}{2}}$$

Those results link Kocher original *DPA* statistics K_k to Legendre-Gauss *Least Squares Method,* to K. Pearson *Product Moment Linear Correlation Coefficient* $\hat{\rho}_{h_k,S}$ and to Fisher *Profile Maximum Likelihood Estimator*. In particular, they show us that the distinction between *partition* (e.g. DPA) and *comparison* (e.g. CPA, MIA) *distinguishers* [81] is an illusion.

Of course,

$$\hat{k} = \arg\max_{k} \hat{\rho}_{h_k,S}{}^2$$

holds as well. But computing $\hat{\rho}_{h_k,S}$ or $\hat{\rho}_{h_k,S}{}^2$ instead of $\left(1-\hat{\rho}_{h_k,S}{}^2\right)^{-\frac{N}{2}}$, as done within CPA attacks, has a serious drawback: it misleads us to the fallacious "Ghost Peaks Problem" [21] that is easily proved to disappear asymptotically with the PMLE.

So, our point is that, if we want to benchmark our new Bayes Factors procedure for the Hamming-Laplace-Gauss *attack model* (and some *non-attack model*), we should better not compare it directly to CPA but rather to Fisher Profile Maximum Likelihood Estimator because K. Pearson correlation and determination coefficients are not directly relevant.

Let us stress also that we should in fact even not try to make such comparison: both procedures arise from completely different theories and reasoning. The Bayes Factors is a post-data procedure, providing a unique solution for each particular data we have, while the PMLE and CPA are pre-data procedures, providing *ad hoc* "solutions" (instead of theorems of Probability Theory) on the average for all possible data that we do not have [40]. Even if those procedures can be compared analytically, the underlying rationale is definitely not the same, rather opposed, and we should clearly not try to compare apples and oranges.

Since this is common practice in the literature [74], let us nevertheless show how we could use *Asymptotic Frequency Sampling Theory* and Signal-to-Noise-Ratio (SNR) reasoning in order to compare them.

On the one hand, for the PMLE/CPA procedure, the signal is

$$\left(1-\hat{\rho}_{h_{k_0},S}{}^2\right)^{-\frac{N}{2}} \qquad k_0 = \arg\max_{k=0,K-1}\left(1-\hat{\rho}_{h_{k_0},S}{}^2\right)^{-\frac{N}{2}}$$

while the background correlation noise at *non-leakage times* is

$$\max_{k=0,K-1}\left(1-\hat{\rho}_{h_k,S}{}^2\right)^{-\frac{N}{2}}$$

because we compute the profile likelihood for each possible value of subkey k. On the other hand, for the Bayes Factors procedure the signal is

$$\hat{\sigma}_{S_{t_j}} \sum_{k=0}^{K-1} \hat{\sigma}_{h_k}{}^{-1}\left(1-\hat{\rho}_{h_k,S}{}^2\right)^{\frac{2-N}{2}} \underset{N\to+\infty}{\simeq} \hat{\sigma}_{S_{t_j}} \hat{\sigma}_{h_{k_0}}{}^{-1}\left(1-\hat{\rho}_{h_{k_0},S}{}^2\right)^{-\frac{N}{2}}$$

So, if we consider that the empirical standard deviations $\hat{\sigma}_{h_k}$ are roughly the same for all possible values of subkey k (due, precisely, to the cryptographic properties of the *decision function*) and that the sample standard deviations $\hat{\sigma}_{S_{i_j}}$ are also roughly the same over time (i.e. homoscedasticity), then the signals are asymptotically the same in both procedures, as we saw in the pictures. But now, the background correlation noise is asymptotically

$$\sum_{k=0}^{K-1}\left(1-\hat{\rho}_{h_k,S}{}^2\right)^{\frac{2-N}{2}} \underset{N\to+\infty}{\simeq} \sum_{k=0}^{K-1}\left(1-\hat{\rho}_{h_k,S}{}^2\right)^{-\frac{N}{2}}$$

So, in order to compare the SNR for both procedures, we should basically compare the moments of the respective frequency sampling distributions of the background correlation noise

$$\max_{k=0,K-1}\left(1-\hat{\rho}_{h_k,S}{}^2\right)^{-\frac{N}{2}} \qquad \text{and} \qquad \sum_{k=0}^{K-1}\left(1-\hat{\rho}_{h_k,S}{}^2\right)^{-\frac{N}{2}}$$

This is interesting from a pure mathematical point of view because we have the *Additive Theory of Random Variables* versus *Extreme Values Theory*.

The asymptotic frequency sampling distribution of the sample correlation coefficients $\hat{\rho}_{h_k,S}$ should be available, for instance in Fisher works. From it, we would get the asymptotic distribution of the profile likelihoods $\left(1-\hat{\rho}_{h_k,S}{}^2\right)^{-\frac{N}{2}}$ by a change of variable. If we neglect their weak mutual dependency and regard those profile likelihoods as i.i.d. random variables over the subkey space, with common p.d.f. $f\left(x\right)$ and c.d.f. $F(x)$, we would get on the one hand the p.d.f. of

$$\sum_{k=0}^{K-1}\left(1-\hat{\rho}_{h_k,S}{}^2\right)^{-\frac{N}{2}}$$

as the K-th convolution power of $f\left(x\right)$ and, on the other hand, the c.d.f. of

$$\max_{k=0,K-1}\left(1-\hat{\rho}_{h_k,S}{}^2\right)^{-\frac{N}{2}}$$

as the K-th power of $F(x)$. Then, we could derive their respective moments. Our experiments and toy simulations make us conjecture that we should find something like

$$E\left[\sum_{k=0}^{K-1}\left(1-\hat{\rho}_{h_k,S}{}^2\right)^{-\frac{N}{2}}\right] \ll E\left[\max_{k=0,K-1}\left(1-\hat{\rho}_{h_k,S}{}^2\right)^{-\frac{N}{2}}\right]$$

and that the inequality should get stronger and stronger as K increases.

From this point of view, our experimental results on the DES with $K = 2^6$ may be again the worst results we can get with the Bayes Factor procedure compared to PMLE/CPA. For example, we would have, in increasing order of relative power: mono DES S-box attacks (6 bits), mono AES S-box attacks (8 bits), double DES S-box attacks (12 bits) and double AES S-box attacks (16 bits).

It would remain to study also the small sample properties of the Bayes Factors, in particular the effect of the strong intrinsic mutual dependence of the *decision functions* $F(m,k), k = 0, K-1$ over the submessages space inducing, by transitivity, a strong mutual dependence of the *profile likelihoods* (i.e. the "Ghost Peaks Problem" for small N [21].) at *leakage times*. At a first glance, the situation could be also favourable to the Bayes Factors procedure, because we sum over significant, non-zero determination coefficients of the same magnitude at *leakage times*, while we consider only maxima in PMLE/CPA attacks. However, this hypothetical effect has not yet been detected in our basic preliminary experiments.

- **Second- and high-order Hamming-Laplace-Gauss "DPA-like" attacks with constant, noisy Gaussian "non-attack" model**.

Consider the second-order *"DPA-like" attack datum*

$$D^i_{t_{j1} \wedge t_{j2}} = m^i \wedge S^i_{t_{j1}} \wedge S^i_{t_{j2}}$$

and its second-order *"DPA-like"* Hamming-Laplace-Gauss *attack datum model*

$$\mathrm{M}^{iH}_{t_{j1} \wedge t_{j2}} = \mathrm{M}^{iH}_{t_{j1}} \wedge \mathrm{M}^{iH}_{t_{j2}} = {}'S^i_{t_{j1}} = \varepsilon_{t_{j1}} h\left[F(m,k) \otimes r^i\right] + \mu_{t_{j1}} + \xi_{t_{j1}}{}' \wedge {}'S^i_{t_{j2}} = \varepsilon_{t_{j2}} h\left(r^i\right) + \mu_{t_{j2}} + \xi_{t_{j2}}{}'$$

with hyperparameters

$$\Theta^{iH}_{t_{j1} \wedge t_{j2}} = \Theta^H_{t_{j1}} \wedge \Theta^H_{t_{j2}} \wedge r^i \qquad \Theta^H_{t_{j1}} = \varepsilon_{t_{j1}} \wedge \mu_{t_{j1}} \wedge \sigma_{t_{j1}} \qquad \Theta^H_{t_{j2}} = \varepsilon_{t_{j2}} \wedge \mu_{t_{j2}} \wedge \sigma_{t_{j2}}$$

Note that we could add some hyperparameters in order to take a possible mutual dependence of $S^i_{t_{j1}}$ and $S^i_{t_{j2}}$ conditionally upon $k \wedge \Theta^{iH}_{t_{j1} \wedge t_{j2}}$ into account (but keep in mind what we said about the independence "assumption".).

The second-order *attack data* at *generalized time* $t_{j_1} \wedge t_{j_2}$ are

$$D^A_{t_{j1} \wedge t_{j2}} = \bigwedge_{i=1}^{N} D^i_{t_{j1} \wedge t_{j2}}$$

and their joint second-order Hamming-Laplace-Gauss *attack data model* is

$$\mathrm{M}^{AH}_{t_{j1} \wedge t_{j2}} = \bigwedge_{i=1}^{N} \mathrm{M}^{iH}_{t_{j1} \wedge t_{j2}}$$

with *joint hyperparameters*

$$\Theta^{H}_{t_{j_1}\wedge t_{j_2}} = \bigwedge_{i=1}^{N} \Theta^{iH}_{t_{j_1}\wedge t_{j_2}} = \Theta^{H}_{t_{j_1}} \wedge \Theta^{H}_{t_{j_2}} \wedge \bigwedge_{i=1}^{N} r^{i}$$

because $A \wedge A = A$ in Boolean Logic. The direct probabilities of exchangeable *"DPA-like" attack data* given the parameters and the models write as, if we regard $S^{i}_{t_{j_1}}$ and $S^{i}_{t_{j_2}}$ as independent conditionally upon parameters $k \wedge \Theta^{i}_{t_{j_1}\wedge t_{j_2}}$ (the conditional independence "assumption".):

$$p\left(D^{A}_{t_{j_1}\wedge t_{j_2}} \middle| k \wedge \Theta^{A}_{t_{j_1}\wedge t_{j_2}} \wedge \mathrm{M}^{A}_{t_{j_1}\wedge t_{j_2}} \wedge I^{A}\right) = \prod_{i=1}^{N} p\left(S^{i}_{t_{j_1}} \wedge S^{i}_{t_{j_2}} \middle| m^{i} \wedge k \wedge \Theta^{A}_{t_{j_1}\wedge t_{j_2}} \wedge \mathrm{M}^{A}_{t_{j_1}\wedge t_{j_2}} \wedge I^{A}\right) p\left(m^{i} \middle| k \wedge \Theta^{A}_{t_{j_1}\wedge t_{j_2}} \wedge \mathrm{M}^{A}_{t_{j_1}\wedge t_{j_2}} \wedge I^{A}\right)$$

$$= \prod_{i=1}^{N} p\left(S^{i}_{t_{j_1}} \middle| m^{i} \wedge k \wedge \Theta^{A}_{t_{j_1}\wedge t_{j_2}} \wedge \mathrm{M}^{A}_{t_{j_1}} \wedge I^{A}\right) p\left(S^{i}_{t_{j_2}} \middle| m^{i} \wedge k \wedge \Theta^{A}_{t_{j_1}\wedge t_{j_2}} \wedge \mathrm{M}^{A}_{t_{j_1}} \wedge I^{A}\right) p\left(m^{i} \middle| k \wedge \Theta^{A}_{t_{j_1}\wedge t_{j_2}} \wedge \mathrm{M}^{A}_{t_{j_1}\wedge t_{j_2}} \wedge I^{A}\right)$$

$$= \prod_{i=1}^{N} p\left(S^{i}_{t_{j_1}} \middle| m^{i} \wedge k \wedge \Theta_{t_{j_1}} \wedge \bigwedge_{u=1}^{N} r^{u} \wedge \mathrm{M}^{A}_{t_{j_1}} \wedge I^{A}\right) p\left(S^{i}_{t_{j_2}} \middle| m^{i} \wedge k \wedge \Theta_{t_{j_2}} \wedge \bigwedge_{u=1}^{N} r^{u} \wedge \mathrm{M}^{A}_{t_{j_2}} \wedge I^{A}\right) p\left(m^{i} \middle| k \wedge \Theta^{A}_{t_{j_1}\wedge t_{j_2}} \wedge \mathrm{M}^{A}_{t_{j_1}\wedge t_{j_2}} \wedge I^{A}\right)$$

$$= \prod_{i=1}^{N} p\left(S^{i}_{t_{j_1}} \middle| m^{i} \wedge k \wedge \Theta_{t_{j_1}} \wedge r^{i} \wedge \mathrm{M}^{i}_{t_{j_1}} \wedge I^{A}\right) p\left(S^{i}_{t_{j_2}} \middle| m^{i} \wedge k \wedge \Theta_{t_{j_2}} \wedge r^{i} \wedge \mathrm{M}^{i}_{t_{j_2}} \wedge I^{A}\right) p\left(m^{i} \middle| k \wedge \Theta^{A}_{t_{j_1}\wedge t_{j_2}} \wedge \mathrm{M}^{A}_{t_{j_1}\wedge t_{j_2}} \wedge I^{A}\right)$$

For model $\mathrm{M}^{AH}_{t_{j_1}\wedge t_{j_2}}$ and uniformly distributed known, non-chosen *submessages*, we have:

$$p\left(D^{A}_{t_{j_1}\wedge t_{j_2}} \middle| k \wedge \Theta^{A}_{t_{j_1}\wedge t_{j_2}} \wedge \mathrm{M}^{AH}_{t_{j_1}\wedge t_{j_2}} \wedge I^{A}\right) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_{t_{j_1}}} e^{-\frac{\left(S^{i}_{t_{j_1}} - \varepsilon_{t_{j_1}} h_{k,r^{i}} - \mu_{t_{j_1}}\right)^{2}}{2\sigma_{t_{j_1}}{}^{2}}} \frac{1}{\sqrt{2\pi}\sigma_{t_{j_2}}} e^{-\frac{\left(S^{i}_{t_{j_2}} - \varepsilon_{t_{j_2}} h_{r^{i}} - \mu_{t_{j_2}}\right)^{2}}{2\sigma_{t_{j_2}}{}^{2}}} M^{-1}$$

$$= \left(2\pi M\right)^{-N} \left(\sigma_{t_{j_1}} \sigma_{t_{j_2}}\right)^{-N} e^{-\frac{1}{2\sigma_{t_{j_1}}{}^{2}} \sum_{i=1}^{N} \left(S^{i}_{t_{j_1}} - \varepsilon_{t_{j_1}} h_{k,r^{i}} - \mu_{t_{j_1}}\right)^{2} - \frac{1}{2\sigma_{t_{j_2}}{}^{2}} \sum_{i=1}^{N} \left(S^{i}_{t_{j_2}} - \varepsilon_{t_{j_2}} h_{r^{i}} - \mu_{t_{j_2}}\right)^{2}}$$

where

$$h_{k,r^{i}} = h\left[F\left(m^{i}, k\right) \otimes r^{i}\right] \qquad \text{and} \qquad h_{r^{i}} = h\left(r^{i}\right)$$

In the same way, for the second-order Gaussian *"non-attack" datum* and *data model*

$$\mathrm{M}^{\neg AG}_{t_{j_1}\wedge t_{j_2}} = \mathrm{M}^{\neg AG}_{t_{j_1}} \wedge \mathrm{M}^{\neg AG}_{t_{j_2}} = 'S^{i}_{t_{j_1}} = \mu_{t_{j_1}} ' + \xi_{t_{j_1}} " \wedge 'S^{i}_{t_{j_2}} = \mu_{t_{j_2}} ' + \xi_{t_{j_2}} "$$

we get the direct probabilities

$$p\left(D^{A}_{t_{j_1}\wedge t_{j_2}} \middle| \mathrm{N}_{t_{j_1}\wedge t_{j_2}} \wedge \mathrm{M}^{\neg AG}_{t_{j_1}\wedge t_{j_2}} \wedge I^{A}\right) = \left(2\pi M\right)^{-N} \left(\sigma_{t_{j_1}} '\sigma_{t_{j_2}} '\right)^{-N} e^{-\frac{1}{2\sigma_{t_{j_1}}{}'^{2}} \sum_{i=1}^{N} \left(S^{i}_{t_{j_1}} - \mu_{t_{j_1}} '\right)^{2} - \frac{1}{2\sigma_{t_{j_2}}{}'^{2}} \sum_{i=1}^{N} \left(S^{i}_{t_2} - \mu_{t_{j_2}} '\right)^{2}}$$

Therefore the Bayes Factors writes as

$$\forall j = 1, J^2 \quad B^{HG}_{t_{j_1} \wedge t_{j_2}} = \frac{\sum_{k=0}^{K-1} \int_{\Theta_{t_{j_1} \wedge t_{j_2}}} p\left(D^A_{t_{j_1} \wedge t_{j_2}} \middle| k \wedge \Theta^H_{t_{j_1} \wedge t_{j_2}} \wedge \mathrm{M}^{AH}_{t_{j_1} \wedge t_{j_2}} \wedge I^A\right) p\left(k \wedge \Theta^H_{t_{j_1} \wedge t_{j_2}} \middle| \mathrm{M}^{AH}_{t_{j_1} \wedge t_{j_2}} \wedge I^A\right) d\Theta_{t_{j_1} \wedge t_{j_2}}}{\int_{\mathrm{N}_{t_{j_1} \wedge t_{j_2}}} p\left(D^A_{t_{j_1} \wedge t_{j_2}} \middle| \mathrm{N}^G_{t_{j_1} \wedge t_{j_2}} \wedge \mathrm{M}^{\neg AG}_{t_{j_1} \wedge t_{j_2}} \wedge I^A\right) p\left(\mathrm{N}^G_{t_{j_1} \wedge t_{j_2}} \middle| \mathrm{M}^{\neg AG}_{t_{j_1} \wedge t_{j_2}} \wedge I^A\right) d\mathrm{N}_{t_{j_1} \wedge t_{j_2}}} =$$

$$\frac{\sum_{k=0}^{k-1} \sum_{r^1=0}^{R-1} \cdots \sum_{r^N=0}^{R-1} \int_{\sigma_{t_{j_2}}} \int_{\mu_{t_{j_2}}} \int_{\varepsilon_{t_{j_2}}} \int_{\sigma_{t_{j_1}}} \int_{\mu_{t_{j_1}}} \int_{\varepsilon_{t_{j_1}}} \left(\sigma_{t_{j_1}} \sigma_{t_{j_2}}\right)^{-N} e^{-\frac{\sum_{i=1}^{N} \left(S^i_{t_{j_1}} - \varepsilon_{t_{j_1}} h_{k,r^i} - \mu_{t_{j_1}}\right)^2}{2\sigma_{t_{j_1}}{}^2} - \frac{\sum_{i=1}^{N} \left(S^i_{t_{j_2}} - \varepsilon_{t_{j_2}} h_{r^i} - \mu_{t_{j_2}}\right)^2}{2\sigma_{t_{j_2}}{}^2}} p\left(k \wedge \Theta^H_{t_{j_1} \wedge t_{j_2}} \middle| \mathrm{M}^{AH}_{t_{j_1} \wedge t_{j_2}} \wedge I^A\right) d\varepsilon_{t_{j_1}} d\mu_{t_{j_1}} d\sigma_{t_{j_1}} d\varepsilon_{t_{j_2}} d\mu_{t_{j_2}} d\sigma_{t_{j_2}}}{\int_{\sigma_{t_2}'} \int_{\mu_{t_2}'} \int_{\sigma_{t_{j_1}}'} \int_{\mu_{t_{j_1}}'} \left(\sigma_{t_{j_1}}' \sigma_{t_{j_2}}'\right)^{-N} e^{-\frac{\sum_{i=1}^{N} \left(S^i_{t_{j_1}} - \mu_{t_{j_1}}'\right)^2}{2\sigma_{t_{j_1}}'^2} - \frac{\sum_{i=1}^{N} \left(S^i_{t_{j_2}} - \mu_{t_{j_2}}'\right)^2}{2\sigma_{t_{j_2}}'^2}} p\left(\mathrm{N}^G_{t_{j_1} \wedge t_{j_2}} \middle| \mathrm{M}^{\neg AG}_{t_{j_1} \wedge t_{j_2}} \wedge I^A\right) d\mu_{t_{j_2}}' d\sigma_{t_{j_2}}' d\mu_{t_{j_2}}' d\sigma_{t_{j_2}}'}$$

From this we get immediately the Bayes Factors for n-th order attacks based and the same Gaussian models. Using operator-like notations such as

$$\left[\prod_{i=1}^{N} \prod_{l=1}^{n-1} \left(\sum_{r^{i,l}=0}^{R-1}\right)\right] \equiv \sum_{r^{1,1}=0}^{R-1} \cdots \sum_{r^{1,n-1}=0}^{R-1} \cdots \sum_{r^{N,1}=0}^{R-1} \cdots \sum_{r^{N,n-1}=0}^{R-1}$$

we can write in “compact” form

$$\forall j = 1, J^n \quad B^{HG}_{t^n_j} =$$

$$= \frac{\sum_{k=0}^{K-1} \left[\prod_{i=1}^{N} \prod_{l=1}^{n-1} \left(\sum_{r^{i,l}=0}^{R-1}\right)\right] \left[\prod_{l=1}^{n} \left(\int_{\sigma_{t_{jl}}} \int_{\mu_{t_{jl}}} \int_{\varepsilon_{t_{jl}}}\right)\right] \prod_{l=1}^{n} \left[\sigma_{t_{jl}}{}^{-N} e^{-\frac{1}{2\sigma_{t_{jl}}{}^2} \sum_{i=1}^{N} \left(S^i_{t_{jl}} - \varepsilon_{t_{jl}} h_{k,r^{i,l}} - \mu_{t_{jl}}\right)^2}\right] p\left(k \wedge \Theta^H_{t^n_j} \middle| \mathrm{M}^{AH}_{t^n_j} \wedge I^A\right) \left[\prod_{l=1}^{n} \left(d\varepsilon_{t_{jl}} d\mu_{t_{jl}} d\sigma_{t_{jl}}\right)\right]}{\left[\prod_{l=1}^{n} \left(\int_{\sigma_{t_{jl}}'} \int_{\mu_{t_{jl}}'}\right)\right] \prod_{l=1}^{n} \left[\sigma_{t_{jl}}{}'^{-N} e^{-\frac{1}{2\sigma_{t_{jl}}'^2} \sum_{i=1}^{N} \left(S^i_{t_{jl}} - \mu_{t_{jl}}'\right)^2}\right] p\left(\mathrm{N}^G_{t^n_j} \middle| \mathrm{M}^{\neg AG}_{t^n_j} \wedge I^A\right) \left[\prod_{l=1}^{n} \left(d\mu_{t_{jl}}' d\sigma_{t_{jl}}'\right)\right]}$$

Generally speaking, nested sums such as

$$\sum_{r^{1,1}=0}^{R-1} \cdots \sum_{r^{1,n-1}=0}^{R-1} \cdots \sum_{r^{N,1}=0}^{R-1} \cdots \sum_{r^{N,n-1}=0}^{R-1}$$

have exponential computational complexity in the sample size N (and also n!). If this were true in the present case, this would be a good point for security managers: those attacks would be absolutely intractable even for small N. But each mask/hyperparameter $r^{i,l}$ enters only into a single side-channel signal and model, not all of them, so that the situation may be quite different. We postpone the analysis of those radically new n-th order attacks for future works.

We hope that those examples are clear enough to let the reader formally derive the Bayes Factors for his own n-th order models, attacks and joint prior probability distributions, in particular for the *profiling phase* of Template Attacks [1][2][22][31][32][80] with *decision functions* such as $F\left(k^i\right)$ or $F\left(m^i, k_0\right)$, etc.

Nonparametric Solution

In the nonparametric framework, we are no longer provided with side-channel *attack* and *"non-attack"* (parametric) models. Even if *Probability as Logic* is fundamentally parametric (due, in particular, to *Integral Representation Theorems* for discrete propositions that ensure *conditional* iidness [27][41][45] and the *Principle of Maximum Entropy* [39][40][42].), so that it is in fact not clear whether we really need to deal with this case or not, let us just point out that we can nevertheless derive nonparametric solutions to **Problem 1** by following Wolf works [85]. Basically, we would just compute other Bayes Factors.

This would enable us to explicit the asymptotic link between the MIA approach [5][30][81], and the present one, which we mentioned in the introduction. But, by definition, we would still need to marginalize the subkey (when k_0 is not known) and to deal with the "*non-attack*" necessary alternative hypothesis.

2) Problem 2

Problem 2 also admits a formal general solution within *Probability as Logic*.

So, assume that L *generalized leakage times* $t_l^n, l = 1, L$ are known or given *a priori*. Then we have, for instance for *"DPA-like" attacks*, the n-th order *joint attack datum*

$$D_J^i = m^i \wedge \bigwedge_{l=1}^{L} S_{t_l^n}^i$$

"Non-attack" models disappear here and we have only an n-th order *attack datum model* for each *datum* D_J^i

$$\mathrm{M}_J^i = \bigwedge_{l=1}^{L} \mathrm{M}_{t_l^n}^{iA}$$

with *joint hyperparameters* Θ_J^i

$$\Theta_J^i = \bigwedge_{l=1}^{L} \Theta_{t_l^n}^i \wedge \ldots$$

Again, we could add some hyperparameters, for instance a covariance matrix Σ in order to take the mutual dependence of the signals $S_{t_l^n}^i$ conditionally upon $k \wedge \Theta_J^i$ into account (still the conditional independence "assumption".).

Then, we have the n-th order *joint attack data*

$$D_J^A = \bigwedge_{i=1}^{N} D_J^i$$

and the *joint attack model*

$$\mathrm{M}_J^A = \overset{N}{\underset{i=1}{\wedge}} \mathrm{M}_J^i$$

with *joint hyperparameters* Θ_J^A

$$\Theta_J^A = \overset{N}{\underset{i=1}{\wedge}} \Theta_J^i$$

The marginal posterior probability mass function for subkey k writes as

$$\begin{aligned} p\left(k \middle| D_J^A \wedge \mathrm{M}_J^A \wedge I^A\right) &= \int_{\Theta_J^A} p\left(k \wedge \Theta_J^A \middle| D_J^A \wedge \mathrm{M}_J^A \wedge I^A\right) d\Theta_J^A \\ &= \int_{\Theta_J^A} \frac{p\left(D_J^A \middle| k \wedge \Theta_J^A \wedge \mathrm{M}_J^A \wedge I^A\right) p\left(k \wedge \Theta_J^A \middle| \mathrm{M}_J^A \wedge I^A\right)}{\sum_{k=0}^{K-1} \int_{\Theta_J^A} p\left(D_J^A \middle| k \wedge \Theta_J^A \wedge \mathrm{M}_J^A \wedge I^A\right) p\left(k \wedge \Theta_J^A \middle| \mathrm{M}_J^A \wedge I^A\right) d\Theta_J^A} d\Theta_J^A \end{aligned}$$

or simply

$$p\left(k \middle| D_J^A \wedge \mathrm{M}_J^A \wedge I^A\right) \propto \int_{\Theta_J^A} p\left(D_J^A \middle| k \wedge \Theta_J^A \wedge \mathrm{M}_J^A \wedge I^A\right) p\left(k \wedge \Theta_J^A \middle| \mathrm{M}_J^A \wedge I^A\right) d\Theta_J^A$$

Choosing as our best candidate

$$\hat{k} = \arg\max_k p\left(k \middle| D_J^A \wedge \mathrm{M}_J^A \wedge I^A\right)$$

we get Aristotle *Maximum a Posteriori Estimator* (MAP) for subkey k_0. This is the unique solution under 0-1 loss function $\delta\left(\hat{k} - k_0\right)$ within *Decision Theory* [40][74] (but k being a discrete parameter, this loss function is unique so that we do not really need Decision Theory.).

A very interesting special case is the following. We could know *a priori* or we could assume that the physical hyperparameters at each *generalized leakage time* $t_l^n, l = 1, L$ are all identical and equal to Θ_0. For n-th order attacks

$$\Theta_J^A = \overset{N}{\underset{i=1}{\wedge}} \overset{L}{\underset{l=1}{\wedge}} \Theta_{t_l^n}^i = \Theta_0 \wedge \overset{N}{\underset{i=1}{\wedge}} \overset{n-1}{\underset{l=1}{\wedge}} r^{i,l}$$

with, for instance,

$$\Theta_0 = \varepsilon_0 \wedge \mu_0 \wedge \sigma_0$$

for the Hamming-Laplace-Gauss *leakage model*. This could happen if the same electronics, the same hardware logic, is involved each time the *decision function(s)* under attack is/are processed by the device. In this case, the problem dimension and the computational burden would be drastically reduced and we would be certainly provided with pretty powerful attacks, if this hypothesis is confirmed *a posteriori*: they would be equivalent to attacks with only one *generalized leakage time* and LN side-channel signals. More generally, in assigning joint prior probability distributions for the physical hyperparameters, we could try to take their intra-model and inter-time mutual logical dependences into account.

Minka [58][59] is an excellent starting point for calculations with multivariate Gaussian direct distributions. See also [11].

III General Solution: Merging Both Sub-Problems

When it is not possible (e.g. sequential attacks with upper bounded sample size N) or when we do not want to solve both sub-problems independently, we must solve them simultaneously, in one shot. After all, this is nothing but the original, genuine DPA/CPA/MIA-like attacks by contrast to Template Attacks or stochastic methods.

But we saw that **Problem 1** belongs to Hypotheses Testing Theory while **Problem 2** belongs to Parameter Estimation Theory. So, how to solve both of them at the same time, within a single theory?

Fortunately, there is in fact a trivial and perfect **duality** between Hypotheses Testing and Parameter Estimation in our system of inference, *Probability as Logic*,... and only in our system of inference [40]:

> Estimating a discrete (resp. a continuous) parameter is nothing but testing a finite (resp. an uncountable infinite) number of hypotheses.

In fact, testing the hypotheses $\left\{H_q, q \in \Omega\right\}$ is nothing but estimating the parameter/index q itself.

While this is really trivial in our system, we would like to take the time to show that it is not at all in non-Bayesian, frequentist, orthodox Statistics [62][63], because it is important to appreciate the simple general solution to come in full extent.

To make it simple, consider the most basic statistical binary hypotheses test that we can find in the security field: the so-called *Monobit Test* for random bits generators [64]. We want to test that the RBG is balanced, i.e. that its binomial proportion $\theta \simeq \Pr\left(bit = 1\right)$ [49] is equal to $\frac{1}{2}$:

$$H_0 : \theta = \frac{1}{2}$$

against the *omnibus*, unspecified alternative

$$H_1 = \neg H_0 : \theta \neq \frac{1}{2}$$

Under the null hypothesis and the iidness of the RBG output bits, the Hamming weight of N such bits follows a binomial $B\left(N, \frac{1}{2}\right)$ distribution. From this we can get easily a *confidence interval* for a given *significance level* α and sample size N. At a first glance, this procedure is very simple and straightforward.

But now, please remember that this kind of tests was introduced by people like K. Pearson, J. Neyman and E. S. Pearson (Karl son) in order to overcome the alleged arbitrariness in the prior probabilities of the hypotheses. According to Neyman [63] (author translation):

We know that the problems of verifying hypotheses were treated by Thomas Bayes. The solutions depended on prior probabilities. Those ones being unknown in general, we were forced to make arbitrary hypotheses about them that rendered the results inapplicable to practical problems.

35 years ago, Karl Pearson published a method for verifying a particular statistical hypothesis, a method known as the χ^2. There was nothing about a priori probabilities in this Memoir that played a remarkable role…

So, basically those tests are supposed to allow us not to use (Bayesian) Probability Theory just because it was, at least more than 70 years ago, incomplete and/or perhaps corrupted by subjectivity. Let us nevertheless examine the Monobit problem from the point of view of *Probability as Logic*. As usual, we would like to compute to posterior probability of H_0 given RBG output data D. By Bayes Rule:

$$p\left(H_0 \middle| D \wedge I\right) = p\left(H_0 \middle| I\right) \frac{p\left(D \middle| H \wedge I\right)}{p\left(D \middle| I\right)}$$

According to Neyman, the potential trouble would lie in the arbitrariness of the prior probability $p\left(H_0 \middle| I\right)$. But elementary Measure Theory tells us (at least the author) that this probability is not arbitrary but in fact exactly equal to 0 (the set $\left\{\frac{1}{2}\right\}$ having Lebesgue measure zero). As a consequence, whatever the data D we may get,

$$p\left(H_0 \middle| D \wedge I\right) = p\left(H_0 \middle| I\right) = 0$$

Now, the elementary Monobit problem itself appears to be at best trivial… unless we are ready to assign non-zero probabilities to negligible sets. So, it is not easy to understand how those who were precisely educated in Measure-theoretic “probability” (i.e. “relative frequency”) [48][60] are nevertheless not reluctant to apply Neyman-Pearson point-null hypotheses testing on continuous parameters instead of relying on (Bayesian) Probability Theory. See [6] for recent and more sophisticated examples in Linear Cryptanalysis (the question is: why don’t the authors even try to follow Shannon, a *reasonable man* according to Maxwell?). Moreover, why should we partition the natural parameter/hypotheses space $[0,1]$ into $\left\{\left\{\frac{1}{2}\right\}, [0,1] \backslash \left\{\frac{1}{2}\right\}\right\}$ and damage it?

But what would be a more meaningful problem, according to *Probability as Logic*? As usual, we would simply estimate parameter θ by computing its (marginal) posterior probability distribution

$$p\left(\theta \middle| D \wedge I\right) = \frac{p\left(D \middle| \theta \wedge I\right) p\left(\theta \middle| I\right)}{\int_0^1 p\left(D \middle| \theta \wedge I\right) p\left(\theta \middle| I\right) d\theta}$$

whose existence is proved for infinitely and finitely exchangeable bits sequences [27][40][41][45], which is precisely a much weaker and much more applicable “assumption” than iidness in the Monobit Test. Then we would take, for instance, the first moment of this distribution (i.e. Laplace *Rule of Succession*) for point estimation under quadratic loss function or derive *Highest Posterior Density Regions* (HPDR) for interval estimation [40][74]. This is a Parameter Estimation problem that we could also interpret, on request even if it is less natural, as testing a *continuum* of hypotheses

$$H_{\lambda} : \theta = \lambda \quad \lambda \in [0,1]$$

Of course, we could also estimate parameter θ without using *Probability as Logic*, for instance as the empirical relative frequency of 1’s if we *select* Fisher popular and standard *Maximum Likelihood* point *Estimator* (but see [50] for a very disturbing consequence of this), or we could derive *confidence intervals*, etc.

Let’s summarize the situation: non-probabilistic methods were introduced in order to overcome some alleged defects of (Bayesian) Probability Theory. As a result, we get two completely different theories: Hypotheses Testing Theory and Estimation Theory. Those theories are so different that we can find elementary problems that are tackled by one theory while they should better be by the other.

In our case, we definitely need to deal with one (many) hypotheses testing problem (**Problem 1**), and one parameter estimation problem (**Problem 2**) simultaneously. Due to the drastic lack of (mutual) logical consistency between orthodox Hypotheses Testing and Parameter Estimation (e.g. Fisher Estimation-theoretic Maximum Likelihood Estimator is mixed up with Neyman-Pearson Hypotheses Testing-theoretic Lemma in [6]. But Neyman and Fisher could not agree on Testing (NP-Lemma versus P-values) as well as on Estimation (frequentist versus fiducial inferences)), we are absolutely unable to see how we could achieve our objective in those frameworks.

Fortunately, we have a single theory for both kinds of problems within *Probability as Logic*. In fact, there is even no need for any special *ad hoc* theory for those problems: we just need to *compute* probabilities and nothing else, as stated by Shannon.

Anyway, in order to solve **Problem 1** and **Problem 2** in one shot, we just have to transform **Problem 1** into a parameter estimation problem. In this way, we will fall into a global, even bigger parameter estimation problem to be solved just as we formally solved **Problem 2** by marginalizing all hyperparameters.

So, let us introduce first new auxiliary hyperparameters $q_{t_j^n} \in \{0,1\}$, with $q_{t_j^n} = 1$ if $H_{t_j^n}^0$ is true and $q_{t_j^n} = 0$ if $H_{t_j^n}^1$ is true in **Problem 1**. By definition, we have:

$$D_{t_j^n}^i \,\Big|\, k \wedge \Theta_{t_j^n}^i \wedge 'q_{t_j^n} = 1' \wedge \mathrm{M}_{t_j^n}^{iA} \wedge I^A \;\sim\; \mathrm{M}_{t_j^n}^{iA}$$

and

$$D_{t_j^n}^i \,\Big|\, k \wedge \mathrm{N}_{t_j^n}^i \wedge 'q_{t_j^n} = 0' \wedge \mathrm{M}_{t_j^n}^{i\neg A} \wedge I^A \;\sim\; \mathrm{M}_{t_j^n}^{i\neg A}$$

or simply

$$D^i_{t^n_j}\Big|k\wedge\Theta^i_{t^n_j}\wedge N^i_{t^n_j}\wedge q_{t^n_j}\wedge\mathrm{M}^{iA}_{t^n_j}\wedge\mathrm{M}^{i\neg A}_{t^n_j}\wedge I^A\ \sim\ q_{t^n_j}\mathrm{M}^{iA}_{t^n_j}+\left(1-q_{t^n_j}\right)\mathrm{M}^{i\neg A}_{t^n_j}$$

if we allow ourselves to identify models and direct probability distributions.

For instance, for first-order *"DPA-like" attacks* with Hamming-Laplace-Gauss *attack datum model* $\mathrm{M}^{AH}_{t_j}$, Gaussian *"non-attack" datum model* $\mathrm{M}^{\neg AG}_{t_j}$ and uniformly distributed submessages, we have

$$p\left[D^i_{t_j}\Big|k\wedge\Theta^H_{t_j}\wedge N^G_{t_j}\wedge q_{t_j}\wedge\mathrm{M}^{AH}_{t_j}\wedge\mathrm{M}^{\neg AG}_{t_j}\wedge I^A\right]=$$

$$p\left[S^i_{t_j}\Big|m^i\wedge k\wedge\Theta^H_{t_j}\wedge N^G_{t_j}\wedge q_{t_j}\wedge\mathrm{M}^{AH}_{t_j}\wedge\mathrm{M}^{\neg AG}_{t_j}\wedge I^A\right]p\left[m^i\Big|k\wedge\Theta^H_{t_j}\wedge N^G_{t_j}\wedge q_{t_j}\wedge\mathrm{M}^{AH}_{t_j}\wedge\mathrm{M}^{\neg AG}_{t_j}\wedge I^A\right]=$$

$$M^{-1}\left[q_{t_j}\left(2\pi\right)^{-\frac{1}{2}}\sigma_{t_j}{}^{-1}e^{-\frac{1}{2\sigma_t{}^2}\left(S^i_{t_j}-\varepsilon_{t_j}h^i_k-\mu_{t_j}\right)^2}+\left(1-q_{t_j}\right)\left(2\pi\right)^{-\frac{1}{2}}\sigma_{t_j}{}'^{-1}\,e^{-\frac{1}{2\sigma_{t_j}{}'^2}\left(S^i_{t_j}-\mu_{t_j}{}'\right)^2}\right]$$

Then, by the *Theorem of Total Probability*, it follows that

$$D^i_{t^n_j}\Big|k\wedge\Theta^i_{t^n_j}\wedge N^i_{t^n_j}\wedge\mathrm{M}^{iA}_{t^n_j}\wedge\mathrm{M}^{i\neg A}_{t^n_j}\wedge I^A\ \sim\ p_{t^n_j}\mathrm{M}^{iA}_{t^n_j}+\left(1-p_{t^n_j}\right)\mathrm{M}^{i\neg A}_{t^n_j}$$

where $p_{t^n_j}=p\left('q_{t^n_j}=1'\Big|k\wedge\Theta^H_{t_j}\wedge N^G_{t_j}\wedge\mathrm{M}^{AH}_{t_j}\wedge\mathrm{M}^{\neg AG}_{t_j}\wedge I^A\right)$.

Thus it appears that, unless *generalized leakage times* are known or given *a priori* (i.e. **Problem 2**. E.g. *extraction phase* of Template Attacks.), original, genuine, elementary (marginal) side-channel *datum* and *data models* **must be** averages between *attack* and *"non-attack" models*. This is known as *Bayesian Model Averaging* [37][74][83]. Since we do not known them *a priori* (e.g. by collusion) most of the time, as in Kocher original DPA [47] and one-shot attacks, we regard this fact, this logical necessity as the *very essence of Side-channel Cryptanalysis*.

In other words, side-channel direct probability distributions/likelihoods **cannot be** frequency sampling distributions. As a consequence it appears that basic, original side-channel problems cannot be tackled at all in frequentist, orthodox Statistics [62][63] because, by definition, they conceive only frequency sampling distributions, but only, *a posteriori*, by (Bayesian) Probability Theory and *Probability as Logic*. As far as we known, this is the first concrete, real-world example of this kind of problems in Cryptanalysis. The problem is that all statistical procedures proposed so far were picked up from non-Bayesian Statistics and signal processing techniques. As a consequence, we know from now on that they miss the very point, which is the logical conjunction of **Problem 1** and **Problem 2**, and that the general solution and attack to come is fundamentally, logically and practically different from any of them.

So, if the present approach is proved to be successful by real-world experiments and more powerful attacks, as we can expect since it is a theorem of Probability Theory, it could have some impact on other areas of Cryptography and Cryptanalysis as we would know that we **should better not** interpret and regard probability distributions as frequency sampling distributions. This would be no more a matter of obscure “philosophy” but of concrete results in practical attacks and security. As a consequence, for the sake of logical consistency, we may like not to rely on the Monobit Test anymore but rather to estimate the RBG binomial proportion or we may like to replace, in χ^2 Cryptanalysis [82], K. Pearson χ^2 test by Jaynes ψ test [40], which is, on the contrary, a theorem of Probability Theory, if we ever need to perform hypotheses tests. And so on and so forth. After all, Side-channel Cryptanalysis may be much more important and fundamental from a conceptual and theoretical point of view than expected.

Equivalently but more generally, we can simply index $Q_{t^n_j}$ mutually exclusive and exhaustive *datum models* and their hyperparameters themselves by $q_{t^n_j}$

$$D^i_{t^n_j} \left| k \wedge \bigwedge_{q_{t^n_j}=0}^{Q_{t^n_j}-1} \Theta^i_{q_{t^n_j}} \wedge q_{t^n_j} \wedge \bigwedge_{q_{t^n_j}=0}^{Q_{t^n_j}-1} \mathrm{M}^i_{q_{t^n_j}} \wedge I^A \right. \sim \mathrm{M}^i_{q_{t^n_j}}$$

For instance, for one *attack datum model* and one *“non-attack” datum model*, we write

$$\mathrm{M}^i_{0_{t^n_j}} = \mathrm{M}^{i\neg A}_{t^n_j} \qquad \text{and} \qquad \mathrm{M}^i_{1_{t^n_j}} = \mathrm{M}^{iA}_{t^n_j}$$

Now, as before, we have the n-th order *attack data*

$$D^A = \bigwedge_{i=1}^{N} \bigwedge_{j=1}^{J^n} D^i_{t^n_j}$$

and the n-th order side-channel *attack and “non-attack” model*

$$\mathrm{M}^{A\neg A} = \bigwedge_{i=1}^{N} \bigwedge_{j=1}^{J^n} \bigwedge_{q_{t^n_j}=0}^{Q_{t^n_j}-1} \mathrm{M}^i_{q_{t^n_j}}$$

with hyperparameters

$$\Theta^{A\neg A} = \bigwedge_{i=1}^{N} \bigwedge_{j=1}^{J^n} \bigwedge_{q_{t^n_j}=0}^{Q_{t^n_j}-1} \Theta^i_{q_{t^n_j}} \wedge \bigwedge_{j=1}^{J^n} q_{t^n_j}$$

As an example, the joint direct probability distribution for n-th order exchangeable *“DPA-like” attack data*

$$D^A = \bigwedge_{i=1}^{N} m^i \wedge \bigwedge_{j=1}^{J^n} S^i_{t^n_j}$$

with uniformly distributed known, non-chosen submessages writes as, if we regard the side-channel signals over *generalized time* as mutually logically independent conditionally upon all parameters (again, remember the conditional independence “assumption”.)

$$p\left(D^A \middle| k\wedge\Theta^{A\neg A}\wedge\mathrm{M}^{A\neg A}\wedge I^A\right)=$$

$$p\left(\bigwedge_{i=1}^{N} m^i \wedge \bigwedge_{j=1}^{J^n} S^i_{t_j^n} \middle| k\wedge\bigwedge_{i=1}^{N}\bigwedge_{j=1}^{J^n}\bigwedge_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\Theta^i_{q_{t_j^n}} \wedge \bigwedge_{j=1}^{J^n} q_{t_j^n}\wedge\mathrm{M}^{A\neg A}\wedge I^A\right)=$$

$$\prod_{i=1}^{N} p\left(m^i \wedge \bigwedge_{j=1}^{J^n} S^i_{t_j^n} \middle| k\wedge\bigwedge_{j=1}^{J^n}\bigwedge_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\Theta^i_{q_{t_j^n}} \wedge \bigwedge_{j=1}^{J^n} q_{t_j^n}\wedge\mathrm{M}^{A\neg A}\wedge I^A\right)=$$

$$\prod_{i=1}^{N} p\left(\bigwedge_{j=1}^{J^n} S^i_{t_j^n} \middle| m^i\wedge k\wedge\bigwedge_{j=1}^{J^n}\bigwedge_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\Theta^i_{q_{t_j^n}} \wedge \bigwedge_{j=1}^{J^n} q_{t_j^n}\wedge\mathrm{M}^{A\neg A}\wedge I^A\right) p\left(m^i \middle| k\wedge\bigwedge_{j=1}^{J^n}\bigwedge_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\Theta^i_{q_{t_j^n}} \wedge \bigwedge_{j=1}^{J^n} q_{t_j^n}\wedge\mathrm{M}^{A\neg A}\wedge I^A\right)=$$

$$M^{-N}\prod_{i=1}^{N}\prod_{j=1}^{J^n} p\left(S^i_{t_j^n} \middle| m^i\wedge k\wedge\bigwedge_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\Theta^i_{q_{t_j^n}} \wedge q_{t_j^n}\wedge\bigwedge_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\mathrm{M}^i_{q_{t_j^n}}\wedge I^A\right)$$

Then, the marginal posterior probability mass function for subkey k writes as, as usual

$$p\left(k \middle| D^A\wedge\mathrm{M}^{A\neg A}\wedge I^A\right)\propto$$

$$\int_{\Theta^{A\neg A}} p\left(D^A \middle| k\wedge\Theta^{A\neg A}\wedge\mathrm{M}^{A\neg A}\wedge I^A\right)p\left(k\wedge\Theta^{A\neg A} \middle| \mathrm{M}^{A\neg A}\wedge I^A\right)d\Theta^{A\neg A}\propto$$

$$\left[\prod_{j=1}^{J^n}\left(\sum_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\right)\right]\left[\prod_{i=1}^{N}\prod_{j=1}^{J^n}\prod_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\left(\int_{\Theta^i_{q_{t_j^n}}}\right)\right]p\left(D^A \middle| k\wedge\bigwedge_{i=1}^{N}\bigwedge_{j=1}^{J^n}\bigwedge_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\Theta^i_{q_{t_j^n}} \wedge \bigwedge_{j=1}^{J^n} q_{t_j^n}\wedge\mathrm{M}^{A\neg A}\wedge I^A\right)\ldots$$

$$\ldots p\left(k\wedge\bigwedge_{i=1}^{N}\bigwedge_{j=1}^{J^n}\bigwedge_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\Theta^i_{q_{t_j^n}} \wedge \bigwedge_{j=1}^{J^n} q_{t_j^n} \middle| \mathrm{M}^{A\neg A}\wedge I^A\right)\left[\prod_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\prod_{j=1}^{J^n}\prod_{i=1}^{N}\left(d\Theta^i_{q_{t_j^n}}\right)\right]\propto$$

$$\left[\prod_{j=1}^{J^n}\left(\sum_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\right)\right]\left[\prod_{i=1}^{N}\prod_{j=1}^{J^n}\prod_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\left(\int_{\Theta^i_{q_{t_j^n}}}\right)\right]\prod_{i=1}^{N}\prod_{j=1}^{J^n}p\left(S^i_{t_j^n} \middle| m^i\wedge k\wedge\bigwedge_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\Theta^i_{q_{t_j^n}} \wedge q_{t_j^n}\wedge\bigwedge_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\mathrm{M}^i_{q_{t_j^n}}\wedge I^A\right)\ldots$$

$$\ldots p\left(k\wedge\bigwedge_{i=1}^{N}\bigwedge_{j=1}^{J^n}\bigwedge_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\Theta^i_{q_{t_j^n}} \wedge \bigwedge_{j=1}^{J^n} q_{t_j^n} \middle| \mathrm{M}^{A\neg A}\wedge I^A\right)\left[\prod_{q_{t_j^n}=0}^{Q_{t_j^n}-1}\prod_{j=1}^{J^n}\prod_{i=1}^{N}\left(d\Theta^i_{q_{t_j^n}}\right)\right]$$

Finally, we are provided with fairly general attacks!

An important special case is the following. We may like to use all *decision functions* $F_d\left(m^i,k\right), d=1,D$ available at the same time, instead of a single one (as we did in **Problem 1**): *multi-decision functions attacks*. How does it work? Those *decision functions* induce the side-channel *attack datum models* $\mathrm{M}^{iA_d}_{t_j^n}, d=1,D$ respectively, with common underlying side-channel *leakage model* $\mathrm{M}^{iL}_{t_j^n}$ and hyperparameters $\Theta^i_{t_j^n}$. If at most one *decision function* is processed at each *generalized time* t_j^n , then we have mutually exclusive and exhaustive hypotheses and the *side-channel models* conditional on index $q_{t_j^n}$ writes as:

$$D_{t_j^n}^i \left| k \wedge \Theta_{t_j^n}^i \wedge N_{t_j^n} \wedge q_{t_j^n} \wedge \bigwedge_{q_{t_j^n}=0}^{D} \mathrm{M}_{q_{t_j^n}}^i \right. \sim \mathrm{M}_{q_{t_j^n}}^i , \quad q_{t_j^n} = 0, D$$

with

$$\mathrm{M}_{0_{t_j^n}}^i = \mathrm{M}_{t_j^n}^{i\neg A} \equiv \mathrm{M}_{t_j^n}^{\neg A} \qquad \text{and} \qquad \forall q_{t_j^n} = 1, D, \quad \mathrm{M}_{q_{t_j^n}}^i = \mathrm{M}_{t_j^n}^{iA_{q_{t_j^n}}}$$

Then, the marginal posterior probability mass function for subkey k reduces to

$$p\left(k \middle| D^A \wedge \mathrm{M}^{A\neg A} \wedge I^A\right) \propto$$

$$\left[\prod_{j=1}^{J^n}\left(\sum_{q_{t_j^n}=0}^{D}\quad\right)\right]\left[\prod_{i=1}^{N}\prod_{j=1}^{J^n}\left(\int_{\Theta_{t_j^n}^i}\quad\right)\right]\left[\prod_{j=1}^{J^n}\left(\int_{\mathrm{N}_{t_j^n}}\quad\right)\right]\prod_{i=1}^{N}\prod_{j=1}^{J^n} p\left(S_{t_j^n}^i \middle| m^i \wedge k \wedge \Theta_{t_j^n}^i \wedge \mathrm{N}_{t_j^n} \wedge q_{t_j^n} \wedge \bigwedge_{q_{t_j^n}=0}^{D} \mathrm{M}_{q_{t_j^n}}^i \wedge I^A\right)...$$

$$...p\left(k \wedge \bigwedge_{i=1}^{N}\bigwedge_{j=1}^{J^n} \Theta_{t_j^n}^i \wedge \bigwedge_{j=1}^{J^n}\left(q_{t_j^n} \wedge \mathrm{N}_{t_j^n}\right) \middle| \mathrm{M}^{A\neg A} \wedge I^A\right)\left[\prod_{j=1}^{J^n}\left(d\mathrm{N}_{t_j^n}\right)\right]\left[\prod_{j=1}^{J^n}\prod_{i=1}^{N}\left(d\Theta_{t_j^n}^i\right)\right]$$

Of course, we could also introduce one *leakage model* for each *decision function* and/or several *"non-attack" models*, or we could use additive total power consumption attack models if several *decision functions* can be processed in parallel: all that is already included in our general analytic formula for the posterior probability mass function for subkey k .

But because we are precisely able from now on to solve **Problem 1** and **Problem 2** simultaneously in a unified logical framework, let us show that this formula is in fact even more general than what we can expect at a first glance.

As an example, consider a basic desynchronisation issue: each side-channel signal $S_{t_j}^i$, $i = 2, N$ is translated by an unknown discrete lag τ^i with respect to the reference signal $S_{t_j}^1$ (this is always more or less the case in practice). The standard way would be to find those lags and to resynchronize those signals (using well-known signal processing techniques such as intercorrelation functions $R_{xy}(\tau)$) before performing the attack itself. But in our approach, we do not need to perform any signal processing by itself: we just have new hyperparameters τ^i, $i = 2, N$ entering into our global problem and we can just write down

$$p\left(k \middle| D^A \wedge \mathrm{M}^{A\neg A} \wedge I^A\right) \propto$$

$$\left[\prod_{i=2}^{N}\left(\sum_{\tau^i=-J}^{J}\quad\right)\right]\int_{\Theta^{A\neg A}} p\left(D^A \middle| k \wedge \bigwedge_{i=2}^{N} \tau^i \wedge \Theta^{A\neg A} \wedge \mathrm{M}^{A\neg A} \wedge I^A\right) p\left(k \wedge \bigwedge_{i=2}^{N} \tau^i \wedge \Theta^{A\neg A} \middle| \mathrm{M}^{A\neg A} \wedge I^A\right) d\Theta^{A\neg A}$$

and perform the calculation with more complex, desynchronized, *contaminated* models. But if we write

$$\forall i = 1, N \quad \Theta^i_{q_{t^n_j}} = \Theta^i_{q_{t^n_j}}{}' \wedge \tau^i \qquad\qquad \tau^1 = 0 \Leftrightarrow p\left(\tau^1 \middle| \mathrm{M}^{A\neg A} \wedge I^A\right) = \delta\left(\tau^1\right)$$

we see that our general formula still holds as it is.

We see no reasons why it would not be the same story with more complex pre-signal processing issues. On the contrary, it would be extremely interesting if we could ever find exceptions to this. In any case, we should be able to do the full job in one shot just by plugging our ignorance and knowledge into the equations. So our formula should encapsulate pre-signal processing issues as well. However, for computational complexity concerns, we acknowledge that we may like to pre-process the signals once and for all! But, for this purpose, it would be nice to use *Probability as Logic* instead of *ad hoc* signal processing methods that may destroy a significant amount of useful information in an uncontrolled way [40].

Anyway, what we have here is something like the hardcore logical structure of generic Side-channel Cryptanalysis, at least on symmetric block ciphers, and we regard

$$p\left(k \middle| D^A \wedge \mathrm{M}^{A\neg A} \wedge I^A\right) \propto$$

$$\left[\prod_{j=1}^{J^n}\left(\sum_{q_{t^n_j}=0}^{Q_{t^n_j}-1}\right)\right]\left[\prod_{i=1}^{N}\prod_{j=1}^{J^n}\prod_{q_{t^n_j}=0}^{Q_{t^n_j}-1}\left(\int_{\Theta^i_{q_{t^n_j}}}\right)\right] p\left(D^A \middle| k \wedge \bigwedge_{i=1}^{N}\bigwedge_{j=1}^{J^n}\bigwedge_{q_{t^n_j}=0}^{Q_{t^n_j}-1}\Theta^i_{q_{t^n_j}} \wedge \bigwedge_{j=1}^{J^n} q_{t^n_j} \wedge \mathrm{M}^{A\neg A} \wedge I^A\right)\ldots$$

$$\ldots p\left(k \wedge \bigwedge_{i=1}^{N}\bigwedge_{j=1}^{J^n}\bigwedge_{q_{t^n_j}=0}^{Q_{t^n_j}-1}\Theta^i_{q_{t^n_j}} \wedge \bigwedge_{j=1}^{J^n} q_{t^n_j} \middle| \mathrm{M}^{A\neg A} \wedge I^A\right)\left[\prod_{q_{t^n_j}=0}^{Q_{t^n_j}-1}\prod_{j=1}^{J^n}\prod_{i=1}^{N}\left(d\Theta^i_{q_{t^n_j}}\right)\right]$$

as its *Fundamental Equation* because it encapsulates (and corrects) all generic parametric attacks known so far, including Correlation Attacks, Template Attacks, high-order attacks, stochastic methods, multi-decision functions attacks, multi-models attacks, signal processing issues, etc., for arbitrary *attack* and *"non-attack" models*, joint prior probability distributions and order. In particular, it should allow us to answer the question raised at the beginning of [79]. The trick was just to solve **Problem 1** and **Problem 2** simultaneously by introducing auxiliary hyperparameters $q_{t^n_j}$.

So, *Probability as Logic* has enabled us to translate many conceptual issues (e.g. designing one more *ad hoc* attack or distinguisher for a new attack scenario) into analytic, symbolic, numerical and computational ones. In engineering words, we can say that we have something like the High Level Design of a nice "SCA (parametric) Machine". For sure there is a long but exciting way before converting this HLD into practical implementations in order to fulfil completely Shannon programme.

Problem 1, **Problem 2** and their logical conjunction being of general interest, we hope that it should be quite straightforward to extend this formal approach, if needed, to other targets, such as stream or asymmetric ciphers and special attacks.

Conclusions

We have undertaken to follow Shannon way to (Side-channel) Cryptanalysis, restarting with known, non-chosen messages attacks on symmetric block ciphers. Since we must deal with probabilities of single and determined cases, first of all we had to place ourselves in a suitable Bayesian system of epistemic probability, logically independent of (mass) (random) events, random variables or relative frequencies. Our favourite one is known as *Probability as Logic*.

This powerful and flexible system has allowed us, in a first step, to get general solutions to two independent generic problems in the parametric framework: the determination of *generalized leakage times* and the determination of the target, given such *generalized leakage times*. The recipe is always the same: write down the joint direct probability distribution conditional on the parameters and the models, assign the joint prior probability distribution of those parameters, apply Bayes Rule to get the joint posterior distribution and marginalize, (dis)integrate out all hyperparameters to get the marginal posterior probability distribution for the *attack data* or the target.

The solution to the first problem was found to be fundamentally different from all approaches proposed so far as we must introduce *"non-attack" models* and as we must marginalize the targets of those attacks. Preliminary experimental results and theoretical analysis indicate that we can expect this new exact procedure to perform well compared to the State-of–the-Art especially for large target spaces, but more works are required in order to quantify this improvement accurately. In addition, we were able to get formal solutions for high-order attacks in a straightforward way, as theorems of *Probability as Logic*, for arbitrary side-channel models.

But the essence of Side-channel Cryptanalysis is precisely to be able to solve both problems simultaneously. This was also easy to achieve in our system of inference and only in our system of inference since we have a single logically consistent theory for both Hypotheses Testing and Parameter Estimation problems.

At this point, it appeared that (marginal) side-channel models are, generally speaking, averages between *attack* and *"non-attack" models* and, more generally, between several conditional models, so that direct distributions/likelihoods cannot be frequency sampling distributions. As a consequence, we do not see, *a priori* and also *a posteriori*, how generic side-channel problems could even be attacked properly in non-Bayesian Hypotheses Testing and Estimation theories because, by definition, they recognize only frequency sampling distributions. As far as we know, this is the first (counter)example of such problems in Cryptanalysis.

If analytic, symbolic, numerical and computational implementation issues can be well managed, we can expect very interesting results and phenomena with this new approach. For instance, it should be possible to guess a target with high probability without knowing the *generalized leakage times* themselves with significant probabilities. But clearly, there is a long but exciting way before going from the HLD to practical implementations of such "SCA Machine" and finally fulfilling Shannon programme.

Essentially, it remains to explain how to get honest discrete side-channel models by logical inspection and how to assign joint prior probability distributions for their parameters, to extend this approach to the nonparametric framework, to apply it to other cryptographic primitives such as stream or asymmetric ciphers and especially to derive formal adaptive chosen-messages attacks. Again, we believe this can be achieved only within this framework.

Ultimately, if this new approach to Side-channel Cryptanalysis is found to be successful, as it *should* be, it could have some impact on completely different areas of Cryptography and Cryptanalysis because we would finally acknowledge that probability distributions should better not be regarded mainly as frequency sampling distributions. This would be no longer a matter of obscure “philosophy” but a matter of practical results in real-world attacks and security. As a consequence, for the sake of logical consistency, we may like to reinterpret and/or to reformulate a couple of concepts accordingly and to replace some statistical procedures by their Bayesian, probabilistic counterparts.

Acknowledgements

This paper is an extremely modest tribute to Professor E.T. Jaynes (1922-1998), *who saw the truth and preserved it*. It is dedicated to my family, especially to the Love of my Life MJ and the baby to come. Special thanks to Pascal Orosco and Patrick Mauss for strong support. In memory of Doctor Antoine Valembois.

Selected References

1. Agrawal D., Rao J. R., Rohatgi P., Schramm K., *Template as Master Keys*, International Workshop on Cryptographic Hardware and Embedded systems N^{o}7, Edinburgh, UK (29/08/2005) 2005, vol. 3659, pp. 15-29
2. Archambeau C., Peeters E., Standaert F.-X., Quisquater J.-J., *Template Attacks in Principal Subspaces,* In L. Goubin and M. Matsui (Eds.), 8th International Workshop on Cryptographic Hardware and Embedded Systems (CHES), Yokohama, Japan, 10-13 October, 2006. Lecture Notes in Computer Science vol. 4249, pp. 1-14. Springer, available at http://www.cs.ucl.ac.uk/staff/C.Archambeau/publ/ches_ca06.pdf
3. Aristotle, *Nicomachean Ethics*, 350 BC, available at http://virtuescience.com/nicomachean-ethics.html
4. Arnborg S., Sjödin G., *What is the plausibility of probability?*, 2003, available at ftp://ftp.nada.kth.se/pub/documents/Theory/Stefan-Arnborg/m2001.pdf
5. Aumônier S., *Generalized Correlation Power Analysis*, Tools for Cryptanalysis Workshop, 2007, available at http://www.impan.gov.pl/BC/Program/conferences/07Crypt-abs/Aumonier%20-%20SubmissionWorkshopSA.pdf
6. Baignières Th., Stern J., Vaudenay S., *Linear Cryptanalysis of Non Binary Ciphers with an Application to SAFER*, available at http://www.di.ens.fr/~stern/data/St122.pdf
7. Bernardo J. M., *Bayesian Statistics*, Encyclopedia of Life Support Systems (EOLSS). Probability and Statistics, (R. Viertl, ed), Oxford, UK: UNESCO, 2003, available at http://www.uv.es/bernardo/BayesStat.pdf
8. Bernoulli Jakob, *Ars Conjectandi sive Stochastica*, 1713
9. Bertrand J., *Calcul des Probabilités*, 1889, available at http://gallica.bnf.fr/ark:/12148/bpt6k99602b/f8
10. Bévan R., *Estimation Statistique et Sécurité des Cartes à Puce - Évaluation d'attaques DPA Evoluées*, Ph. D Thesis, Université de Paris-Sud 11, Orsay, June 2004
11. Bishop Ch. M., Tipping M., *Bayesian Regression and Classification*, Advances in Learning Theory: Methods, Models and Applications, J.A.K. Suykens et al. (Editors), IOS Press, NATO Science Series III: Computer and Systems Sciences, volume 190, 2003, available at http://research.microsoft.com/~cmbishop/downloads/Bishop-NATO-Bayes.pdf
12. Borel E., *Valeur Pratique et Philosophie des Probabilités*, Jacques Gabay, 1939
13. Borel E., *Le Hasard*, Presses Universitaires de France, 1948
14. Borel E., *Une Objection à la Définition Empirique de la Probabilité*, Comptes rendus hebdomadaires des séances de l'Académie des sciences, *T.211 (1940) 312-313*
15. Berger J. O., Liseo B, Wolpert R. L, *Integrated Likelihood Methods for Eliminating Nuisance Parameters*, *Statistical Science*, Vol. 14, No. 1 (Feb., 1999), pp. 1-22, available at http://citeseer.ist.psu.edu/berger97integrated.html
16. Bernardo J. M. (2003), *Bayesian Statistics*, *Encyclopedia of Life Support Systems (EOLSS). Probability and Statistics*, (R. Viertl, ed.). Oxford, UK: UNESCO, available at http://www.uv.es/~bernardo/BayesStat2.pdf
17. Bretthorst G. L., *An Introduction to Parameter Estimation using Bayesian Probability Theory,* in Maximum Entropy and Bayesian Methods, Dartmouth, 1989, pp. 53-79, P. F. Fougère (ed.), Kluwer Academic Publishers, The Netherlands, 1990, available at http://citeseer.ist.psu.edu/bretthorst90introduction.html
18. Bretthorst G. L., *An introduction to Model Selection using Probability as Logic*, in *Maximum Entropy and Bayesian Methods*, G. R. Heibreder (ed.), pp. 1-42, 1996
19. Bretthorst G. L., *The Near-Irrelevance of Sampling Frequency Distributions,* in *Maximum Entropy and Bayesian Methods,* W. von der Linden *et al.* (eds.)*,* pp. 21-46, Kluwer Academic Publishers, Dordrecht the Netherlands, 1999
20. Bretthorst G. L., *Bayesian Spectrum Analysis and Parameter Estimation*, Lecture Notes in Statistics, 48, 1988, Springer Verlag, ISBN 0-387-96871-7, ISBN 3-540-96871-7, available at http://bayes.wustl.edu/glb/book.pdf
21. Brier E., Clavier Ch., Olivier F., *Correlation Power Analysis with a Leakage Model*, *Lecture Notes in Computer Science 3156,* CHES '04 proceedings*,* pp.16-29, Springer Verlag, 2004

22. Chari S., Rao J. R., Rohatgi P., *Template Attacks*, Lecture Notes in Computer Science, Volume 2523/2003, Cryptographic Hardware and Embedded Systems - CHES 2002, Springer Berlin/Heidelberg
23. Coron J.-S., Kocher P., Naccache D., *Statistics and Secret Leakage*, *ACM Transactions on Embedded Computing Systems (TECS),* Volume 3*,* Issue 3*,* pp. 492-508*,* ACM Press*,* August 2004
24. Cox, R. T., *Probability, Frequency and Reasonable Expectation*, American Journal of Physics, 14, 1-13., 1946
25. Cox R. T., *The Algebra of Probable Inference*, The Johns Hopkins Press, 1961
26. Dupré M. J, *Unkowns and Guessed Values. The Laws of Expectation and Probability*, available at http://www.math.tulane.edu/~dupre/shrtguess.pdf
27. de Finetti B., *La Prévision, ses Lois Logiques, ses Sources Subjectives*, Annales de l'institut Henri Poincaré, 7 no. 1 (1937), p. 1-68, available at http://www.numdam.org/item?id=AIHP_1937__7_1_1_0
28. Garrett A.J.M., *Whence the Laws of Probability?*, Maximum Entropy and Bayesian Methods, Proceedings of the 17th International Workshop on Maximum Entropy and Bayesian Methods of Statistical Analysis, held in Boise, Idaho, 1997. Edited by Gary J. Erickson, Joshua T. Rychert, and C. Ray Smith. Kluwer Academic Publishers, Dordrecht/Boston
29. Gibbs J. W., *Elementary Principles in Statistical Mechanics*, Yale University Press, New Haven, Connecticut, 1902
30. Gierlichs B., Batina L., Tuyls P., *Mutual Information Analysis -- A Universal Differential Side-Channel Attack*, available at http://eprint.iacr.org/2007/198
31. Gierlichs B., Lemke-Rust K., Paar Ch., *Templates vs. Stochastic Methods – A Performance Analysis for Side-Channel Cryptanalysis*, International Workshop on Cryptographic Hardware and Embedded Systems N^{o}8, Yokohama , Japan (2006) , vol. 4249, pp. 15-29, available at www.iacr.org/archive/ches2006/02/02.ps
32. Giraud Ch, *Attaques de Cryptosystèmes Embarqués et Contre-mesures Associées*, Ph. D. Thesis, available at http://www.prism.uvsq.fr/fileadmin/CRYPTO/TheseCG-new.pdf
33. Good I. J., *46656 Varieties of Bayesians* (#765), available at http://socrates.berkeley.edu/~fitelson/148/good_bayes.pdf
34. Goldman S., Rivest R. L., *Making Maximum Entropy Constraints Easier By Adding Extra Constraints (Extended Abstract)*, in Maximum--Entropy and Bayesian Methods in Science and Engineering (Vol. 2), (Edited by G.J. Erickson and C.R. Smith) (Kluwer Academic Publishers, 1988), 323—340, available at http://people.csail.mit.edu/rivest/publications.html
35. Goubin L., Patarin J., *DES and Differential Power Analysis*, *The "Duplication" Method, Lecture Notes in Computer Science 1717,* CHES '99 proceedings*,* pp.158-172, Springer Verlag, 1999
36. Hardy M., *Scaled Boolean Algebras*, 2002, available at http://www.stats.org.uk/cox-theorems/Hardy2002.pdf
37. Hoeting J. A., Madigan D., Raftery A. E., Volinsky Ch. T., *Bayesian Model Averaging: a Tutorial*, Statistical Science 14, 382-401, 1999, available at http://www.stat.colostate.edu/~jah/papers/statsci.pdf
38. van Horn K. S., *Constructing a Logic of Plausible Inference: a Guide to Cox's Theorem*, 2003, available at http://www.stats.org.uk/cox-theorems/VanHorn2003.pdf
39. Jaynes E. T., *Information Theory and Statistical Mechanics*, *Phys. Rev.,* Vol. 106**,** 620-630, 1957, available at http://bayes.wustl.edu/etj/articles/theory.1.pdf
40. Jaynes E. T., *Probability Theory: the Logic of Science*, Cambridge University Press, 2003, ISBN-13: 9780521592710, ISBN-10: 0521592712
41. Jaynes, E. T., *Some Applications and Extensions of the de Finetti Representation Theorem*, 1986, in *Bayesian Inference and Decision Techniques with Applications: Essays in Honor of Bruno de Finetti,* P. K. Goel and A. Zellner (eds.), North-Holland, Amsterdam, p. 31, available at http://bayes.wustl.edu/etj/articles/applications.pdf
42. Jaynes, E. T*., Concentration of Distributions at Entropy Maxima,* 1979, *in* E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, R. D. Rosenkrantz (ed.), D. Reidel, Dordrecht, p. 315*
43. Jeffreys H., *Theory of Probability*, Oxford University Press, USA; 3 edition (November 12, 1998), ISBN-10: 0198503687, ISBN-13: 978-0198503682, 1961

44. Joye M, Paillier P., Schoenmakers B., *On Second-order Differential Power Analysis*, CHES 2005, Lecture Notes in Computer Science vol. 3659, pp. 293-308. Springer, disponible sur www.win.tue.nl/~berry/papers/ches05hodpa.pdf
45. Kerns G. J., *Signed Measures in Exchangeability and Infinite Divisibility*, Ph. D. thesis, available at http://cc.ysu.edu/~gjkerns/pdf/Jdiss.pdf
46. Keynes J. M., *A Treatise on Probability*, Rough Draft Printing (June 19, 2008), ISBN-10: 1603861181, ISBN-13: 978-1603861182, 1921
47. Kocher P., Jaffe J., Jun B., *Introduction to Differential Power Analysis and Related Attacks*, available at http://www.cryptography.com/resources/whitepapers/DPATechInfo.pdf, 1998
48. Kolmogorov A. N., *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, Berlin, 1933
49. Kolmogorov A. N., *On Logical Foundations of Probability Theory*, Lecture notes in mathematics, 1983, vol. 1021, pp. 1-5
50. Lad F., Deely J., Piesse A., *Using the Fundamental Theorem of Prevision to Identify Coherency Conditions for Finite Exchangeable Inference*, 1993, available at http://www.math.canterbury.ac.nz/research/rpt95.pdf
51. Laplace P.-S., *Théorie Analytique des Probabilités*, third edition, Gauthier-Villars, 1820
52. Laplace P.-S., *Mémoire sur la Probabilité des Causes par les Événements*, *Savants Étrangers* 6:621-656, also *Œuvres* 8:27-65, 1774
53. Leibniz G.W., *Nouveaux Essais sur l'Entendement Humain*, Flammarion, ISBN 2-08-070582-2, 1704/1764
54. Lewis C., Garnett W., *James Clerk Maxwell, with a selection from his correspondence and occasional writings and a sketch of his contributions to science*, 1882, available at http://www.sonnetsoftware.com/bio/maxbio.pdf
55. Lhoste E., *Le Calcul des probabilités appliqué à l'artillerie, lois de probabilité a priori*, Revue d'artillerie, Mai-Août, Berger-Levrault, Paris, 1923
56. Månsson A., Porta Mana P. G. L., Björk, *`Plausibilities of plausibilities': an approach through circumstances*, 2006, available at http://arxiv.org/abs/quant-ph/0607111/
57. Messerges Th., *Using Second-Order Power Analysis to Attack DPA Resistant Software*, Lecture Notes in Computer Science 1965, 2000, pp. 27-78, Springer Berlin / Heidelberg
58. Minka Th., *Bayesian Linear Regression*, MIT Media Lab Note, 7/19/00, available at http://research.microsoft.com/~minka/papers/linear.html
59. Minka Th., *Inferring a Gaussian Distribution*, MIT Media Lab Note, 1998, available at http://research.microsoft.com/~minka/papers/gaussian.html
60. von Mises R., *Théorie des Probabilités. Fondements et applications*, Annales de l'institut Henri Poincaré, 3 no. 3 (1933), p. 345-345, available at http://www.numdam.org/numdam-bin/fitem?id=AIHP_1933__3_3_345_0
61. de Moivre A., *The Doctrine of Chances: or, a Method of Calculating the Probability of Events in Play*, 1756, available at http://www.ibiblio.org/chance/
62. Mselati B., Benaych-Georges F., *Introduction aux Probabilités et aux Statistiques*, available at http://www.cmapx.polytechnique.fr/~benaych/Probas.03.10.04.pdf
63. Neyman J., *Sur la Vérification des Hypothèses Statistiques Composées*, Bulletin de la S.M.F., tome 63 (1935), p. 246-266, available at http://archive.numdam.org/ARCHIVE/BSMF/BSMF_1935__63_/BSMF_1935__63__246_0/BSMF_1935__63__246_0.pdf
64. NIST Special Publication 800-22, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, 2001, available at http://csrc.nist.gov/publications/nistpubs/800-22/sp-800-22-051501.pdf
65. Oswald E., Mangard S., Herbst C. and Tillich S., *Practical Second-Order DPA Attacks for Masked Smart Card Implementations of Block Ciphers*, Proceedings of CT-RSA 2006, LNCS 3860, Springer, pp. 192-207, available at www.iaik.tugraz.at/Research/sca-lab/publications/pdf/Oswald2006PracticalSecond-OrderDPA.pdf
66. Pascal B., *Oeuvres Complètes*, l'Intégrale/Seuil, ISBN 2-02-000713-4
67. Peeters E., Standaert F.-X., Quisquater J.-J., *Power and Electromagnetic Analysis: Improved Model, Consequences and Comparisons*, Integration, the VLSI Journal, Volume 40 , Issue 1 (January 2007), pp. 56-60, available at www.dice.ucl.ac.be/crypto/files/publications/pdf252.pdf
68. Poincaré H., *Calcul des Probabilités*, second edition, Gauthier-Villars, 1912

69. Pólya G., *Mathematics and Plausible Reasoning*, Princeton University Press (August 3, 1990), ISBN-10: 0691025096, ISBN-13: 978-0691025094, 1954
70. Porta Mana P. G. L., *Studies in plausibility theory, with applications to physics*, PhD thesis, 2007, available at http://www.tp.umu.se/~mana/mana070106-thesis.pdf
71. Porta Mana P. G. L., Månsson A., Björk G.: *The Laplace-Jaynes approach to induction*, 2007, available at http://arxiv.org/abs/physics/0703126/
72. Quisquater J.-J., Samyde D., *Electromagnetic analysis (EMA): Measures and counter-measures for smard cards*, in Proceedings of the International Conference on Research in Smart Cards: Smart Card Programming and Security (E-smart), I. Attali and T. Jensen, Eds., 2001, vol. 2140 of LNCS, pp. 200–210, Springer-Verlag
73. Rissanen J., *A Universal Prior for Integers and Estimation by Minimum Description Length*, The Annals of Statistics, 1983, No. 2, 416-431, available at http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.aos/1176346150
74. Robert Ch. P., *The Bayesian Choice: from Decision-Theoretic Foundations to Computational Implementation*, second edition, Springer, ISBN-10: 0387952314, ISBN-13: 978-0387952314, 2001
75. Rodriguez C., *A Geometric Theory of Ignorance*, available at http://omega.albany.edu:8008/ignorance/
76. Schindler W., Lemke K., Paar Ch., *A Stochastic Model for Differential Side Channel Cryptanalysis*, in *Proceedings of CHES 2005*, LNCS 3659, pages 30--46, Edinburgh, Scotland 2005, available at http://www.iacr.org/archive/ches2005/003.pdf
77. Shafer G., *Comments on "Constructing a Logic of Plausible Inference: A Guide to Cox's Theorem", by Kevin S. Van Horn*, 2003, available at http://www.glennshafer.com/assets/downloads/other14.pdf
78. Shannon C.E., *Communication Theory of Secrecy Systems*, 1949, available at http://netlab.cs.ucla.edu/wiki/files/shannon1949.pdf
79. Standaert F.-X., Malkin T. G., Yung M., *A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks,* available at http://eprint.iacr.org/2006/139.pdf
80. Standaert F.-X., Archambeau C., *Using Subspace-Based Template Attacks to Compare and Combine Power and Electromagnetic Information Leakages*, in E. Oswald and P. Rohatgi (Eds.), 10th International Workshop on Cryptographic Hardware and Embedded Systems (CHES), Washington, DC, USA, 10-13 August, 2008. Lecture Notes in Computer Science vol. 5154, pp. 411-425. Springer, available at http://www.cs.ucl.ac.uk/staff/C.Archambeau/publ/ches_fx08.pdf
81. Standaert F.-X., Gierlichs B., Verbauwhede I., *Partition vs. Comparison Side-channel Distinguishers, An Empirical Evaluation of Statistical Tests for Univariate Side-Channel Attacks against Two Unprotected CMOS Devices*, to appear in the proceedings of ICISC 2008, available at http://www.dice.ucl.ac.be/~fstandae/PUBLIS/60.pdf
82. Vaudenay S., *An Experiment on DES Statistical Cryptanalysis*, Proceedings of the 3rd ACM Conferences on Computer Security, New Delhi, India, pp. 139-147, ACM Press, 1996, available at http://lasecwww.epfl.ch/php_code/publications/search.php?ref=Vau96b
83. Wasserman L., *Bayesian Model Selection and Model Averaging*, available at http://www.stat.cmu.edu/tr/tr666/tr666.html
84. Wikipedia, *Point-biserial Correlation coefficient*, http://en.wikipedia.org/wiki/Point-biserial_correlation_coefficient
85. Wolf D., *Mutual Information as a Bayesian Measure of Independence*, 1995, available at http://arxiv.org/abs/comp-gas/9511002