Redefining the Transparency Order

Kaushik Chakraborty[1], Sumanta Sarkar[2], Subhamoy Maitra[2], Bodhisatwa Mazumdar[3], Debdeep Mukhopadhyay[3], and Emmanuel Prouff[4]

[1] SECRET Team, INRIA, Rocquencourt
kaushik.chakraborty@inria.fr
[2] Indian Statistical Institute, Kolkata
sumanta.sarkar@gmail.com, subho@isical.ac.in
[3] Indian Institute of Technology, Kharagpur
bm.iitkgp@gmail.com, debdeep.mukhopadhyay@gmail.com
[4] Agence nationale de la sécurité des systèmes d'information (ANSSI)
e.prouff@gmail.com

Abstract. In this paper, we consider the multi-bit Differential Power Analysis (DPA) in the Hamming weight model. In this regard, we revisit the definition of Transparency Order (TO) from the work of Prouff (FSE 2005) and find that the definition has certain limitations. Although this work has been quite well referred in the literature, surprisingly, these limitations remained unexplored for almost a decade. We analyse the definition from scratch, modify it and finally provide a definition with better insight that can theoretically capture DPA in Hamming weight model for hardware implementation with precharge logic. At the end, we confront the notion of (revised) transparency order with attack simulations in order to study to what extent the low transparency order of an s-box impacts the efficiency of a side channel attack against its processing. To the best of our knowledge, this is the first time that such a critical analysis is conducted (even considering the original notion of Prouff). It practically confirms that the transparency order is indeed related to the resistance of the s-box against side-channel attacks, but it also shows that it is not sufficient alone to directly achieve a satisfying level of security. Regarding this point, our conclusion is that the (revised) transparency order is a valuable criterion to consider when designing a cryptographic algorithm, and even if it does not preclude to also use classical countermeasures like masking or shuffling, it enables to improve their effectiveness.

Keywords: AES, Auto-correlation, Cross-correlation, Differential Power Analysis, PRINCE, s-box, Transparency Order, Walsh Spectrum.

1 Introduction

Differential Power Analysis (DPA) is one of the strongest forms of side-channel attacks in which the information about the secret key is leaked through power traces while the encryption is being executed on a cryptographic platform. The efficiency of these attacks is naturally much higher than linear or differential cryptanalysis due to the information related to the power traces. To resist such attacks, algorithmic countermeasures like masking [8] and leakage resistant logic [30] exist, that may lead to increased footprint on the implementation platforms in terms of area and power consumptions. Because of phenomenon like glitches, it should be noted that in practical scenarios even masked circuits can be subjected to DPA. With this backdrop, it is evident that the s-boxes in block ciphers would be the prime target of DPA. From the designers point of view, the s-boxes should be chosen carefully such that they should have high DPA resilience in addition to the resistance to other classical cryptanalytic attacks like linear and differential cryptanalysis.

In [15], the theoretical resistance of AES and DES s-boxes to linear cryptanalysis vis a vis DPA attacks in terms of signal-to-noise ratio (SNR) was investigated. Then, an attempt to quantify the DPA resilience of the s-boxes was made in [24], where the parameter Transparency Order (TO) was introduced. This was an important attempt in defining a metric for the DPA resilience of S-boxes for almost a decade ago. Based on a side-channel efficiency metric close to the *standard score* measure involved in [15, 31], the paper [24] tried to explain that s-boxes with smaller TO have higher DPA resilience. The TO, as defined in [24] was found to depend on the propagation characteristics (PC) of the co-ordinate functions of the s-boxes. The bent functions that satisfy the PC for all orders have been found to have worst TO value (though we show in this paper that by the definition of [24], TO cannot be measured for a bent function), while the linear s-boxes have the best DPA resilience. However, the linear s-boxes are not acceptable as a secure cryptographic primitive. Further analyses of TO, as defined in [24], have been followed in *e.g.,* [5, 11, 18, 22].

In this paper, we exhibit several inconsistencies in the original definition given in [24] and we provide an improved definition of the transparency order that appears to be a better metric for quantifying the resistance the resistance of an s-box to DPA attacks. Eventually, its soundness to quantify the resistance of an s-box against side-channel attacks is investigated thanks to several attack simulations. We also provide Appendices A, B, C and D for some detailed calculations, bounds, examples and comparisons with other related works.

Before presenting these contributions, we hereafter start by introducing some useful basics on Boolean functions and DPA attacks which are subsequently used to present our analyses in a formal way.

2 Preliminaries

2.1 Basics of Boolean functions

Let $\mathbb{F}_2^n$ be the vector space that contains all the n-bit binary vectors. For a vector $u \in \mathbb{F}_2^n$, we denote by $\mathsf{H}(u)$ the number of 1's in its binary representation (it is usually referred to as the *Hamming weight* of u). A (single output) Boolean function on n variables may be viewed as a mapping from $\mathbb{F}_2^n$ into $\mathbb{F}_2$. We will denote the set of n-variable Boolean functions as $\mathcal{B}_n$. It is easy to note that $|\mathcal{B}_n| = 2^{2^n}$.

The support of a Boolean function f is defined as $Supp(f) = \{x \in \mathbb{F}_2^n | f(x) = 1\}$. When we use a Boolean function as a cryptographic primitive, we generally consider the functions which output 0 and 1 with equal probability. Thus, we generally consider functions in $\mathcal{B}_n$ for which the cardinality of the support is 2^{n-1}. These are known as *balanced* functions.

Let $x = (x_1, \ldots, x_n)$ and $\omega = (\omega_1, \ldots, \omega_n)$ both belong to $\mathbb{F}_2^n$ and the inner product $x \cdot \omega = x_1\omega_1 \oplus \cdots \oplus x_n\omega_n$. The *Walsh transform* of $f(x)$ is an integer valued function over $\mathbb{F}_2^n$ which is defined as $W_f(\omega) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) \oplus x \cdot \omega}$. The *autocorrelation transform* of $f(x)$ is again an integer valued function over $\mathbb{F}_2^n$ which is defined as $\mathcal{A}_f(\omega) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) \oplus f(x \oplus \omega)}$. The Walsh and autocorrelation spectra are important properties in designing Boolean functions that may be used as cryptographic primitives. In general, it is expected that the maximum absolute value in any of these spectra should be low for better resistance against cryptanalysis (see *e.g.,* [6]). In this paper, we also use the notion of *cross-correlation spectrum* between two Boolean functions; for $f_1, f_2 \in \mathcal{B}_n$, it is defined for every $\omega \in \mathbb{F}_2^n$ as the value $\mathcal{C}_{f_1,f_2}(\omega) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f_1(x) \oplus f_2(x \oplus \omega)}$ (note that we have $\mathcal{C}_{f,f}(\omega) = \mathcal{A}_f(\omega)$).

An $n \times m$ s-box F can be seen as a multi-output Boolean function, namely a function from $\mathbb{F}_2^n$ into $\mathbb{F}_2^m$ with $m \leq n$. Let $u \in \mathbb{F}_2^m$ be a vector whose binary coordinates are all zero except one which is assumed to be at index j. The jth *component function* of the s-box F is the single output Boolean function $u \cdot F$. If $F = (F_1, \ldots, F_m)$, then one may note that $u \cdot F = F_j$. By abuse of notation, we may also denote this component function as F_u. That is, the notations F_u and F_j are used interchangeably for a component function in this paper. The set of linear (resp. affine) functions from $\mathbb{F}_2^n$ into itself will be denoted by $\mathbf{L}_n$ (resp. $\mathbf{A}_n$).

2.2 Basics on DPA attacks

Differential Power Analysis (DPA) is one kind of side channel attack that exploits the difference between the power consumed by a single gate when its output changes from zero to one or vice versa. Initial results in this direction have been presented by Kocher et al [16]. It consists in observing the processing of a cryptographic algorithm (*e.g.,* a block cipher) and in measuring a sample of power consumption traces $\boldsymbol{T}_x$ related to a sufficiently large number of plaintexts x and a constant secret parameter. These traces (which may be viewed as real-valued vectors) actually give the information about the power consumed by each gate when output changes. Next, based on these power traces, the actual attack can be mounted off-line.

In a *single-bit* DPA attack against a block cipher, a particular bit of the intermediate state during the processing is considered and the attacker builds a distinguisher by partitioning the power traces in two bins by predicting whether the bit value is zero or one, corresponding to the key guessed. Let j denote the index of the targeted bit and let $\dot{K}$ denote a secret sub-part that statistically depends on this bit (assuming that the block cipher is iterative, $\dot{K}$ may correspond to the secret parameter of an s-box and is typically 4, 6 or 8 bit long). The attacker makes a guess K on $\dot{K}$ and then the traces are assigned in either of the two bins, say S_0 and S_1, according to hypotheses on the targeted bit which are deduced from K and the plaintexts x. To discriminate the good guess from the wrong ones, [16] proposes to compute the *differential trace* $\boldsymbol{D}_{K,j}$ defined by $\boldsymbol{D}_{K,j} = \frac{1}{|S_1|}\sum_{\boldsymbol{T}_x \in S_1} \boldsymbol{T}_x - \frac{1}{|S_0|}\sum_{\boldsymbol{T}_x \in S_0} \boldsymbol{T}_x$. The quantity $\boldsymbol{D}_{K,j}$ works as a distinguisher in the single-bit power attack model. According to the theory proposed in [16] the vector $\boldsymbol{D}_{K,j}$ should show a peak for the correct key $K = \dot{K}$.

In single-bit `DPA` attack, the sets S_0 and S_1 are constructed on the basis of a fixed bit coordinate j. To improve the efficiency of the attack and to better discriminate the wrong key-hypotheses, a natural idea is to simultaneously consider several bit indices j. This kind of `DPA` attack, initially introduced by Messerges in [20], is called *multi-bit DPA*. It works as follows; for each of the guessed key K, the quantity $\boldsymbol{D}_{K,j}$ is computed for several j (*e.g.*, $j \in [1..m]$) and the results are added to form a new distinguisher $\boldsymbol{D}_K$. As in the single-bit case, $\boldsymbol{D}_K$ is expected to show a peak if $K = \dot{K}$. In [10], it has been proved that this attack is equivalent to the so-called `CPA` attack introduced in [4] up to a change of the attacker model[5]. In [1], it has been proposed to add the absolute values of the $\boldsymbol{D}_{K,j}$ (instead of the values themselves) to build $\boldsymbol{D}_K$. This approach may be a valuable alternative in practice and it is not equivalent to a `CPA` [10].

The work presented in this paper focuses on multi-bit `DPA` which is today systematically tested against industrial cryptographic implementations when it comes to test their security with respect to side-channel attacks. Before proceeding further, let us introduce hereafter a few basic concepts related to Boolean functions.

2.3 `DPA` Attack and Transparency Order of s-boxes

As discussed in Section 1, the `DPA` provides the attacker a verifier (or distinguisher) to guess the correct key. The distinguisher based on the so called *differential traces* $(\boldsymbol{D}_{k,j})_{k \in \mathbb{F}_2^n, j \in [1..m]}$ (resp. $(\boldsymbol{D}_k)_{k \in \mathbb{F}_2^n}$ for multi-bit `DPA`/`CPA`) works on the hypothesis that it takes the maximum value for the correct key (though, due to the presence of system noises one can get several other "ghost peaks" for wrong keys as observed in [4]). Now, from the designer's point of view, one should design the s-boxes in such a manner so that the distribution of the differential trace becomes almost uniform. In such a case, the correct key cannot be distinguished from a wrong key. This idea has been the starting point of two studies published in 2004 [15, 24] aiming at defining a new security criterion for the design of s-boxes with improved resistance against `DPA` attacks.

[5] Such a model is used in `CPA`, together with the key hypothesis K, to compute the predictions that are correlated to each point of the traces $\boldsymbol{T}_x$ (see [4] for a detailed presentation of the `CPA`).

In [15], a few ideas were presented to measure the efficiency of DPA on an s-box. They studied single-bit DPA during the manipulation of the s-box output, the information being assumed to leak in the Hamming distance model with independent additive noise [4]. This model assumes that the leakage takes the form $\mathsf{H}(\beta \oplus F(x \oplus \dot{K})) + B$, where x and $\dot{K}$ respectively denote a plaintext and a round key sub-part, where β denotes the initial content of the register before updating with $F(x \oplus \dot{K})$ and where B denotes an independent (measurement) noise. In this model, the authors of [15] show that the coordinate with highest amplitude in the single-bit distinguisher $\boldsymbol{D}_{K,j}$ related to the j-th coordinate of F is asymptotically equivalent to:

$$\Delta_{K,\dot{K}}(j,\beta) \doteq 2^{-n} \sum_{i=1}^{m} (-1)^{\beta_i} \sum_{x \in \mathbb{F}_2^n} \mathcal{C}_{F_i,F_j}(K \oplus \dot{K}) \ , \tag{1}$$

where β_i denotes the i-th binary coordinate of β.

The original idea of [15] has been afterwards developed and extended in [24] to encompass multi-bit DPA. This leads the author to introduce the notion of *transparency order* (TO for short) to quantify the resistance of s-boxes towards DPA attacks; if some s-box shows low TO, then that s-box is more resistant against DPA attack, *i.e.,* the number of power traces to identify the correct key will be higher. The TO notion introduced in [24] not only depends on the s-box's algebraic properties but also on the register initial state $\beta \in \mathbb{F}_2^m$ which is assumed to be constant for some platforms like smart cards which are based on *precharge logic*. The precharge logic is applied for some microcontrollers in which there is a precharge phase during which the registers are initialized to some fixed value (*e.g.,* before each round in a block cipher implementation). After certain assumptions, the final formula defining the TO of a $n \times m$ s-box $F = (F_1, \ldots, F_m)$ is given by:

$$\mathsf{TO}(F) = \max_{\beta \in \mathbb{F}_2^m} \left(|m - 2\mathsf{H}(\beta)| - \frac{1}{2^{2n} - 2^n} \sum_{a \in \mathbb{F}_2^{n*}} |\sum_{i=1}^{m} (-1)^{\beta_i} \mathcal{A}_{F_i}(a)| \right) . \tag{2}$$

Remark 1. The TO notion share the same basic ideas with the side-channel efficiency metric (*standard score*) discussed in [31]. In those papers, the effectiveness of an attack is measured by computing the difference between the score of the distinguisher for the good key and the average score for the wrong hypotheses, the difference being normalized with the variance of the scores.

One can easily check that the TO becomes minimum for the s-boxes for which the co-ordinate Boolean functions become constant or affine and on the other hand, the highly nonlinear s-boxes have higher transparency order, implying that they are more susceptible to DPA attacks. Though the linear s-boxes are good in terms of TO, linear s-boxes cannot be used for cryptographic reasons and thus we have to mainly study the behaviour of TO for the well known nonlinear s-boxes used in practice. The TO values of constant or linear s-boxes are actually of little interest, though those values may be computed for noting the bounds on TO.

In this paper, we critically study the above definition of TO. In the next section (Section 3), we actually show that this definition has redundancy in terms of considering the maximum over all $\beta \in \mathbb{F}_2^m$. Then in Section 4, we critically analyse the definition of TO in the Hamming weight model from the basic background and identify all the limitations of the existing definition [24]. Next, in Section 5, we present our modified definition of TO that explains the `DPA` in Hamming weight model more appropriately than [24] (see also Remark 5). Eventually, Section 6 confronts the notion of (revised) transparency order with attack simulations. The goal is to investigate how the low transparency order of an s-box impacts the efficiency of a side channel attack against its processing.

3 Redundant Definition of Transparency Order [24]

In this section, we explain why the definition (2) is redundant. For such a purpose, we denote by τ_F^{β} the value $|m-2\mathsf{H}(\beta)|-\frac{1}{2^{2n}-2^n}\sum_{a\in\mathbb{F}_2^{n*}}|\sum_{i=1}^{m}(-1)^{\beta_i}\mathcal{A}_{F_i}(a)|$ in (2) and by $\nu_{F,\beta}$ the value $\frac{1}{2^{2n}-2^n}\sum_{\alpha\in\mathbb{F}_2^{n*}}|\sum_{i=1}^{m}(-1)^{\beta_i}\mathcal{A}_{F_i}(\alpha)|$. With these new notations, we have $\mathsf{TO}(F) = \max_{\beta\in\mathbb{F}_2^m}\tau_{F,\beta}$. We give hereafter our first result.

Proposition 1. $\tau_F^{\beta} = \tau_F^{\overline{\beta}}$.

Proof. Note that $|m-2\mathsf{H}(\beta)| = |m-2(m-\mathsf{H}(\overline{\beta}))| = |-m+2\mathsf{H}(\overline{\beta})| = |m-2\mathsf{H}(\overline{\beta})|$. One may also check that $\nu_{F,\beta} = \nu_{F,\overline{\beta}}$, which directly concludes the proof. $\square$

Next we present the most important result of this section.

Proposition 2. *Let* $0 < \mathsf{H}(\beta) \le \lfloor \frac{m}{2} \rfloor$. *Then* $\tau_F^{\beta} \le \tau_F^{\mathbf{0}}$.

Proof. Consider that $0 < k = \mathsf{H}(\beta) \le \lfloor \frac{m}{2} \rfloor$. Now, $\tau_F^{\mathbf{0}} = (m - \nu_{F,\mathbf{0}})$ and $\tau_F^{\beta} = (m - 2k - \nu_{F,\beta})$.

Let, in contrast to the statement of the proposition, $\tau_F^{\beta} > \tau_F^{\mathbf{0}}$. Then $\nu_{F,\mathbf{0}} - \nu_{F,\beta} > 2k$, *i.e.*,

$$\sum_{\alpha\in\mathbb{F}_2^{n*}}\left[|\sum_{i=1}^{m}(\mathcal{A}_{F_i}(\alpha))| - |\sum_{i=1}^{m}\left((-1)^{\beta_i}\mathcal{A}_{F_i}(\alpha)\right)|\right] > (2^{2n}-2^n)2k.$$

Let $S = \{1, 2, \ldots, m\}$ and $T \subseteq S$, such that $i \in T$ if and only if $\beta_i = 1$. That is T is the support of β.

Then we can rewrite the above inequality as

$$\sum_{\alpha\in\mathbb{F}_2^{n*}}\left[|\sum_{i=1}^{m}(\mathcal{A}_{F_i}(\alpha))| - |\sum_{i\in S\setminus T}(\mathcal{A}_{F_i}(\alpha)) - \sum_{i\in T}(\mathcal{A}_{F_i}(\alpha))|\right] > (2^{2n}-2^n)2k.$$

Using the inequality $|x| - |y| \leq |x - y|$, we obtain,

$$\sum_{\alpha \in \mathbb{F}_2^{n*}} \left[|\sum_{i=1}^{m} (\mathcal{A}_{F_i}(\alpha)) - \sum_{i \in S \setminus T} (\mathcal{A}_{F_i}(\alpha)) + \sum_{i \in T} (\mathcal{A}_{F_i}(\alpha)) | \right] > (2^{2n} - 2^n)2k,$$

$$i.e., \quad \sum_{\alpha \in \mathbb{F}_2^{n*}} 2 \left| \sum_{i \in T} (\mathcal{A}_{F_i}(\alpha)) \right| > (2^{2n} - 2^n)2k.$$

We know that $|\mathcal{A}_{F_i}(\alpha)| \leq 2^n$, and thus we land into a contradiction as the left hand side is always less than or equal to the right hand side. (Even taking the maximum value 2^n, we get that the left hand side is equal, but cannot be greater than the right hand side.) Thus the proof. $\square$

Therefore, we have the following result that shows that the definition of transparency order is actually redundant and it does not depend on β. The proof follows from Propositions 1, 2.

Theorem 1. $\mathsf{TO}(F) = \tau_F^{\mathbf{0}} = m - \frac{1}{2^{2n}-2^n} \sum_{\alpha \in \mathbb{F}_2^{n*}} | \sum_{i=1}^{m} \mathcal{A}_{F_i}(\alpha)|$.

4 Critically analyzing TO for Multi-bit DPA Attack

Given the redundancy in the definition of TO as in [24], we look into the definition from the basic principle and obtain various other limitations of the definition. We highlight several assumptions considered in [24] and critically comment on those.

The output of the s-box becomes $F(x \oplus \dot{K})$ from β, where β is the precharge logic value that is fixed with the system, *i.e.,* β is constant. So, the number of bits, changed after storing the s-box output bits is $\mathsf{H}(F(x \oplus \dot{K}) \oplus \beta)$. The basic idea of DPA works as follows.

Given the correct key $\dot{K}$ (the attacker does not know this), the power trace of the encryption or decryption can be collected by the adversary corresponding to the known plaintexts. As the corresponding $n \times m$ s-boxes are in general not very large, one may expect that the power traces are available corresponding to all the 2^n possible inputs x. Once the data is available, the attack may work off-line where the attacker tries each of the possible 2^n keys K (these are actually parts of round keys that are XORed with x) and partitions the power traces in two bins. It is expected that for the correct key $\dot{K}$, the partitioning will show some distinguishing feature than that in the case of the incorrect keys.

Let us concentrate on the j-th output bit of the s-box. Given any key K (which may or may not be $\dot{K}$), we put the power related information in two bins depending on the value of $F_j(x \oplus K)$. As in [24], for theoretical analysis, the Hamming weight of $F(x \oplus K) \oplus \beta$ can be considered as a logical model for the power related information. The difference of average value in the two bins is

$$\Delta_{K,\dot{K}}(j, \beta) = \frac{1}{|S_{K,1}|} \sum_{x \in S_{K,1}} \mathsf{H}\left(F(x \oplus \dot{K}) \oplus \beta\right) - \frac{1}{|S_{K,0}|} \sum_{x \in S_{K,0}} \mathsf{H}\left(F(x \oplus \dot{K}) \oplus \beta\right),$$

where $S_{K,0} = \{x|F_j(x \oplus K) = 0\}$ and $S_{K,1} = \{x|F_j(x \oplus K) = 1\}$. At this point let us present a technical result.

Proposition 3. *Let $F = (F_1, \ldots, F_m)$ be an $n \times m$ s-box (*i.e., *$F : \mathbb{F}_2^n \to \mathbb{F}_2^m$). Let the co-ordinate functions be such that they are pairwise complement when m is even. In case m is odd, then we consider $\frac{m-1}{2}$ pairs of pairwise complement functions and one constant function. For any such s-box, $\Delta_{K,\dot{K}}(j, 0) = 0$ for any $j,\ 1 \le j \le m$.*

Proof. The proof follows by noting that $\mathsf{H}(F(x))$ is constant for all $x \in \mathbb{F}_2^n$. Thus it is immediate to note that $\mathsf{H}\left(F(x \oplus \dot{K})\right)$ is always constant making $\Delta_{K,\dot{K}}(j, 0) = 0$, for any j. □

Consider a special case when m is even. For any such s-box, where all the functions are taken to be pairwise complement and bent, no `DPA` is possible under the model we are working on. However, the measure presented in [24, 5] shows that TO is maximum (*i.e.,* equals m) for such functions, which means that they are maximally prone to such `DPA`. The conclusion in [24, 5] is not correct and it happened due to certain assumptions that make the definition of TO invalid for s-boxes with unbalanced co-ordinate functions such as bent.

Assumption 1. The analysis of [24] implicitly assumes that the co-ordinate functions of the s-box are balanced.

Note that, while the above assumption that F_j is balanced is true for most popular s-boxes in block ciphers, it is not true in general. Most importantly, it is incorrect to hence apply the old formulation of Transparency Order to bent functions, as done in [24, Theorem 1] and later in [5]. To clarify the situation, we show more explicitly how the assumption of the balancedness of the co-ordinate functions led to the old definition of transparency order.

If F_j is balanced, then $|S_{K,0}| = |S_{K,1}| = 2^{n-1}$ and one can show (see [24] or Equations (17) and (18) in Appendix B for details):

$$\Delta_{K,\dot{K}}(j, \beta) = \frac{1}{2^n} \left((-1)^{\beta_j} \mathcal{A}_{F_j}(K \oplus \dot{K}) + \sum_{i=1, i \neq j}^{m} (-1)^{\beta_i} \mathcal{C}_{F_i,F_j}(K \oplus \dot{K}) \right) \quad , \qquad (3)$$

which implies the following for $K = \dot{K}$;

$$\Delta_{\dot{K},\dot{K}}(j, \beta) = (-1)^{\beta_j} + \frac{1}{2^n} \sum_{i=1, i \neq j}^{m} (-1)^{\beta_i} \mathcal{C}_{F_i,F_j}(0) \ . \qquad (4)$$

Assumption 2 [24]. $F_i \oplus F_j$ is balanced for every $(i, j) \in [1..m]^2$ with $i \neq j$.

Under Assumption 2, we have $\mathcal{C}_{F_i,F_j}(0) = 0$ for every pair of distinct indices i and j and thus, $\Delta_{\dot{K},\dot{K}}(j, \beta) = (-1)^{\beta_j}$. One may note that Assumptions 1 and

2 are satisfied when the involved (n, m) s-box is balanced (which is the case of the vast majority of cryptographic s-boxes).

In single-bit DPA attack the expression $\Delta_{K,\dot{K}}(j,\beta)$ is calculated for a fixed index j and for all hypotheses $K \in \mathbb{F}_2^n$. In the multi-bit case, the latter calculation is done for every $j \in [1..m]$. This actually leads to the processing of the following quantity $\delta_{K,\dot{K}}(\beta)$ [24, Equation (10)]:

$$\delta_{K,\dot{K}}(\beta) = |\sum_{j=1}^{m} \Delta_{K,\dot{K}}(j,\beta)|. \tag{5}$$

For $K \neq \dot{K}$, we have:

$$\delta_{K,\dot{K}}(\beta) = \frac{1}{2^n} \left| \sum_{j=1}^{m} \sum_{i=1}^{m} (-1)^{\beta_i} \mathcal{C}_{F_i,F_j}(K \oplus \dot{K}) \right| . \tag{6}$$

And, for $K = \dot{K}$, we have:

$$\delta_{\dot{K},\dot{K}}(\beta) = \left| \sum_{j=1}^{m} (-1)^{\beta_j} \left(1 + \frac{1}{2^n} \sum_{i=1, i\neq j}^{m} (-1)^{\beta_i \oplus \beta_j} \mathcal{C}_{F_i,F_j}(0) \right) \right| , \tag{7}$$

and Assumption 2 implies $\delta_{\dot{K},\dot{K}}(\beta) = \left| \sum_{j=1}^{m} (-1)^{\beta_j} \right| = |m - 2\mathsf{H}(\beta)|$.

We now introduce the last assumption made in [24].

Assumption 3. The cross-correlation terms $\mathcal{C}_{F_i,F_j}(K \oplus \dot{K})$ can be considered to be zero for every $i \neq j$ and every $(K, \dot{K})$.

Remark 2. Clearly, this third assumption is not true as we cannot have all the values zero in cross-correlation spectrum in general. In fact, we will later present a detailed study for the 8×8 s-box used in AES that will clearly show the limitation of the Assumption 3 made in [24].

Given Assumption 3, we have $\delta_{K,\dot{K}}(\beta) = \frac{1}{2^n} \left| \sum_{j=1}^{m} (-1)^{\beta_j} \mathcal{A}_{F_j}(K \oplus \dot{K}) \right|$. It is expected that if $K = \dot{K}$, then $\delta_{K,\dot{K}}(\beta)$ takes the maximum value. Hence, from the designer's point of view, s-boxes should be chosen in such a manner so that for most of $K \neq \dot{K}$, the value of $\delta_{K,\dot{K}}(\beta)$ does not deviate much from $\delta_{\dot{K},\dot{K}}(\beta)$. The core idea is the following one: a small average distance between $\delta_{K,\dot{K}}(\beta)$ and $\delta_{\dot{K},\dot{K}}(\beta)$ will render the discrimination of $\dot{K}$ more difficult. This led the author of [24] to introduce the following definition of transparency order TO:

$$\mathsf{TO}(F) = \max_{\beta \in \mathbb{F}_2^m} \tau_F^\beta, \text{ where } \tau_F^\beta = \frac{1}{2^n - 1} \sum_{K \in F_2^n - \{\dot{K}\}} \left(\delta_{\dot{K},\dot{K}}(\beta) - \delta_{K,\dot{K}}(\beta) \right) , \tag{8}$$

As we have described, the work in [24] considered a few assumptions, but several critical comments can be made on that. To be specific, we point out the following.

- Assumption 1 is logical and it works for practical s-boxes that the co-ordinate Boolean functions are balanced. However, given this assumption, it is not possible to consider unbalanced co-ordinate functions in the s-box under this definition. However, this had been improperly done in [24, 5] in concluding certain results related to bent functions.
- Assumption 2 is also logical in the sense that it is quite practical to consider that the XOR of two co-ordinate functions should be balanced.
- Assumption 3 considers that the cross-correlation spectrum between two co-ordinate functions will contain all zero values due to certain independence. This is indeed not correct and one should specifically calculate these value that we will also present here. For example, for the 8×8 s-box used in AES, the values in the spectrum are indeed significant. Ignoring this significance led to the redundant definition of $\mathsf{TO}(F)$ in [24] where it does not depend on β at all as we have described in Section 3.

4.1 Considering the cross-correlation terms

For a proper measure, we need to add the cross-correlation terms as given in (6). Thus, we should consider the following definition for further investigation in this area:

$$\begin{aligned} {\tau'}_F^{\beta} &= |m - 2\mathsf{H}(\beta)| \\ &\quad - \frac{1}{2^n(2^n-1)} \sum_{a \in \mathbb{F}_2^{n*}} \left| \sum_{j=1}^{m} \left((-1)^{\beta_j} \mathcal{A}_{F_j}(a) + \sum_{i=1, i \neq j}^{m} (-1)^{\beta_i} \mathcal{C}_{F_i,F_j}(a) \right) \right| . \end{aligned} \tag{9}$$

Similar to Proposition 1, one may note that under the modified definition of ${\tau'}_F^{\beta}$ as given in (9), ${\tau'}_F^{\beta}$ and ${\tau'}_F^{\overline{\beta}}$ are equal (see (19) in Appendix B for details). Let us now consider ${\tau'}_F$ as in (10). This definition will only be valid for the $n \times m$ s-boxes $F = (F_1, \ldots, F_m)$, where each co-ordinate function F_i and further the functions $F_i \oplus F_j$ for $0 \leq i \neq j \leq m$ are balanced.

$$\begin{aligned} {\tau'}_F &= \max_{\beta \in \mathbb{F}_2^m} [|m - 2\mathsf{H}(\beta)| \\ &\quad - \frac{1}{2^{2n} - 2^n} \sum_{a \in \mathbb{F}_2^{n*}} \left| \sum_{j=1}^{m} \left((-1)^{\beta_j} \mathcal{A}_{F_j}(a) + \sum_{i=1, i \neq j}^{m} (-1)^{\beta_i} \mathcal{C}_{F_i,F_j}(a) \right) \right|]. \end{aligned} \tag{10}$$

Unfortunately, similar to (2), this definition also becomes redundant for cryptographically strong s-boxes. This is because, the magnitude of the values in auto-correlation and cross-correlation spectra of cryptographically strong n-variable Boolean functions are of the order of $2^{\frac{n}{2}}$. Thus, the term $|m - 2\mathsf{H}(\beta)|$ will dominate hugely and naturally it will be maximum when β has all-zero or all-one pattern.

Now consider the definition from cryptanalysts' viewpoint. The power traces will be available to the attacker and she can analyse the data off-line to guess the correct key. Thus, the attacker will try to use the power traces in such a way so that the correct key $\dot{K}$ can be distinguished from the other keys.

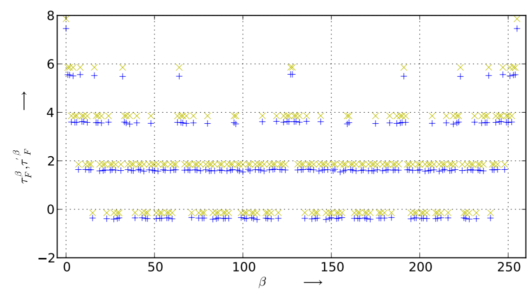


Fig. 1. Values of τ_F^β (in Yellow) and ${\tau'}_F^\beta$ (in Blue) as β varies over $\mathbb{F}_2^8$ for `AES` s-box. The 8-bit patterns of β are written as their integer values.

For `AES` s-box, the values of τ_F^β and ${\tau'}_F^\beta$ are plotted in Figure 1 for different values β. The figure shows that for some values of β, τ_F^β and $\tau_F^{\overline{\beta}}$ are negative. For those values of β, the coefficient $\delta_{K,\dot{K}}(\beta)$, which prevents the attacker to guess the correct key for these β. This happens because the absolute value is processed after taking the sum of the $\Delta_{K,\dot{K}}(j,\beta)$ (see (5)). Note that if the precharge logic β is indeed all-zero or all-one, then the existing idea of [24] works well. However, when $\mathsf{H}(\beta)$ becomes closer to $\frac{m}{2}$, the term $|m - 2\mathsf{H}(\beta)|$ reduces substantially, making τ_F^β and ${\tau'}_F^\beta$ negative in some cases, that makes the definition unacceptable from cryptanalyst's point of view.

5 Redefining TO: where to take the absolute values

As explained before Proposition 3, the quantity $\Delta_{K,\dot{K}}(j,\beta)$ corresponds to the test of the key candidate K in a single-bit `DPA` attack. The attack is assumed to target the j-th bit of the updating with $F(x \oplus \dot{K})$ of a register whose initial content is always β. In [24], the author suggests to extend this attack to several bits by processing the absolute value of the quantities $\Delta_{K,\dot{K}}(j,\beta)$ for several indices j (*e.g.,* $j \in [1..8]$ for an 8-bit register). In [10], it is argued that this kind of attack is equivalent to a `CPA` up to a change of the attacker leakage modeling. Here, we propose another approach to extend the initial single-bit attack to a multi-bit one. This exactly corresponds to the proposal made by Bévan and Knudsen in [1], and consists in summing the absolute values of $\Delta_{K,\dot{K}}(j,\beta)$ for several indices j. It has been proved that this kind of multi-bit `DPA` attack is not reducible to a CPA attack and is actually a valuable alternative in practice [10]. Our proposal leads to the following definition of a coefficient $\underline{\delta}_{K,\dot{K}}(\beta)$:

$$\underline{\delta}_{K,\dot{K}}(\beta) = \sum_{j=1}^{m} |\Delta_{K,\dot{K}}(j,\beta)| \ . \tag{11}$$

Similarly as for (7), Assumption 2 implies that $\underline{\delta}_{\dot{K},\dot{K}}(\beta)$ equals $\sum_{j=1}^{m} |(-1)^{\beta_j}|$ that is m. Compared to (7), it may be checked that the term related to $\mathsf{H}(\beta)$ is removed. This is indeed important since this Hamming weight of β was influencing the complete measure without any real justification and was making the definition redundant by maximizing it for all-zero or all-one β.

For $K \neq \dot{K}$, Equation (6) becomes:

$$\begin{aligned} \underline{\delta}_{K,\dot{K}}(\beta) &= \frac{1}{2^n} \sum_{j=1}^{m} \left| (-1)^{\beta_j} \mathcal{A}_{F_j}(K \oplus \dot{K}) + \sum_{i=1, i\neq j}^{m} (-1)^{\beta_i} \mathcal{C}_{F_i,F_j}(K \oplus \dot{K}) \right| \\ &= \frac{1}{2^n} \sum_{j=1}^{m} \left| \sum_{i=1}^{m} (-1)^{\beta_i \oplus \beta_j} \mathcal{C}_{F_i,F_j}(K \oplus \dot{K}) \right| \ , \end{aligned} \tag{12}$$

which leads to the following new version, called $\mathsf{TO}(F,\beta)$, of the coefficient defined in (9); $\mathsf{TO}(F,\beta) \doteq \frac{1}{2^n-1} \sum_{a \in \mathbb{F}_2^{n\star}} \underline{\delta}_{0,0}(\beta) - \underline{\delta}_{a,0}(\beta)$, or equivalently

$$\mathsf{TO}(F,\beta) = \frac{1}{2^n - 1} \sum_{a \in \mathbb{F}_2^{n\star}} \left(m - \sum_{j=1}^{m} |\Delta_{a,0}(j,\beta)| \right) \tag{13}$$

that is

$$\mathsf{TO}(F,\beta) = m - \frac{1}{2^n(2^n-1)} \sum_{a \in \mathbb{F}_2^{n*}} \sum_{j=1}^{m} \left| \sum_{i=1}^{m} (-1)^{\beta_i \oplus \beta_j} \mathcal{C}_{F_i,F_j}(a) \right| \ , \tag{14}$$

where a plays the role of $K \oplus \dot{K}$ in (12).

Remark 3. We still have $\mathsf{TO}(F,\beta) = \mathsf{TO}(F,\overline{\beta})$.

We eventually deduce the following new definition of the TO criterion:

Definition 1 (Improved Transparency Order). *Let F be a balanced $n \times m$ function. Its improved transparency order is the coefficient $\mathsf{TO}(F)$ defined by:*

$$\mathsf{TO}(F) = \max_{\beta \in \mathbb{F}_2^m} \left(m - \frac{1}{2^{2n} - 2^n} \sum_{a \in \mathbb{F}_2^{n*}} \sum_{j=1}^{m} \left| \sum_{i=1}^{m} (-1)^{\beta_i \oplus \beta_j} \mathcal{C}_{F_i,F_j}(a) \right| \right) \ . \tag{15}$$

Remark 4. In order not to multiply the number of terms and criteria, we chose to continue to use the term "transparency order" when referring to our improved version. As we think that the previous version of the criterion introduced in [24] should be systematically replaced by the new one, this should not introduce ambiguity in the rest of this work.

Remark 5. From designers' point of view, it is important to consider the precharge values β for which $\mathsf{TO}(F,\beta)$ is minimum. Indeed, for such precharge values β, the attacker advantage is minimum. In the rest of the paper, the value $\min_{\beta\in\mathbb{F}_2^m} \mathsf{TO}(F,\beta)$ is denoted by $\mathsf{TO}_{\min}(F)$ and called *minimum transparency order*. On the other hand, precharge logic values maximizing $\mathsf{TO}(F,\beta)$ correspond to the worst case from designers' point of view and to the best case for attackers.

In Appendix C, we exhibit an interesting lower bound on $\mathsf{TO}(F)$. The main combinatorial contribution in this bound is that, all the cross-correlation terms are replaced by Walsh spectrum values.

5.1 Example with some existing s-boxes

Let us first refer to the 8×8 s-box F of `AES` [13]. The graphical representation $\mathsf{TO}(F,\beta)$ is presented in Figure 2 and one may note the symmetry due to $\mathsf{TO}(F,\beta) = \mathsf{TO}(F,\overline{\beta})$.

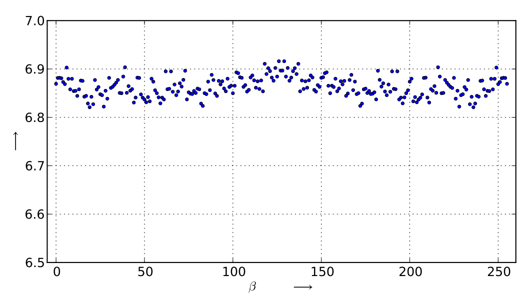


Fig. 2. Values of $\mathsf{TO}(F,\beta)$ as β varies over $\mathbb{F}_2^8$ for AES s-box. The 8-bit patterns of β are written as their integer values.

It may be observed that the minimum value 6.82083 of $\mathsf{TO}(F,\beta)$ is achieved for $\beta = \beta_{\min} = (0,0,0,1,0,0,1,1)$ (integer value 19) and its complement, whereas the maximum value 6.91605 is achieved for $\beta_{\max} = (0,1,1,1,1,1,1,0)$ (integer value 126) and its complement. We hence deduce that the TO of the `AES` s-box is $\mathsf{TO}(F) = 6.91605$. Moreover, from designers' point of view, it is better to use $\beta_{\min}$ as the precharge logic as in that case more effort will be required (than when any other β is the precharge logic) to identify the peak corresponding to the correct key. Our lower bound of $\mathsf{TO}(F)$ presented in Theorem 4 of Appendix C provides the value 6.51457. We have also analyzed the family of `PRINCE` 4×4 s-boxes [3] (details in Appendix A). Our analysis shows that they have significantly different behaviours in terms of TO. This is experimentally confirmed in Section 6.

5.2 TO of s-boxes in the same (extended) affine equivalence classes

Two $n \times n$ s-boxes F and G are affine equivalent if there exist two affine permutations $A, B \in \mathbf{A}_n$ such that $G = B \circ F \circ A$, *i.e.,* $G(x) = [B \circ F \circ A](x)$, for all $x \in \mathbb{F}_2^n$. Further if there exists $C \in \mathbf{L}_n$ such that F and G satisfy $G(x) = [B \circ F \circ A](x) \oplus C(x)$, then F and G are said to be *extended affine* (EA) equivalent. Some cryptographic properties remain invariant under such transformations. The set of functions G which are EA-equivalent to F is denoted by EA(F). In this section, we would like to pursue whether the DPA resistance of s-boxes is affected because of these transformations. It is an important aspect in design of block ciphers, which often assumes that since affine transformation does not affect properties like non-linearity, maximum value in the auto-correlation spectrum, degree (when more than 1), it is fine to replace one with the other.

The affine invariance of the TO criterion is stated in the following theorem whose proof is given in Appendix B.

Theorem 2. *Let F be an $n \times m$ balanced function. Then, for any affine permutation $A \in \mathbf{A}_n$ we have* $\mathsf{TO}(F \circ A) = \mathsf{TO}(F)$.

On the other hand, if two functions F and G are EA-equivalent but not affine equivalent (meaning that $G = B \circ F \circ A \oplus C$ with $A, B \in \mathbf{A}_n$, $C \in \mathbf{L}_n$ and C non-constant), then their TO are not necessarily equal. For example, $F =$ 084b2ef613ac57d9 and $G =$ 03d5070601ebeff0 are two EA-equivalent s-boxes with $\mathsf{TO}(F) = 2.466$ and $\mathsf{TO}(G) = 2.766$. These two s-boxes are related by the relation $F = B \circ F + C$, where B and C are the linear mappings associated with the matrices $\begin{bmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 1&0&1&0 \\ 0&1&0&1 \end{bmatrix}$ and $\begin{bmatrix} 0&1&0&1 \\ 1&0&1&0 \\ 1&0&0&1 \\ 1&0&1&1 \end{bmatrix}$, respectively.

The number of 4×4 s-boxes up to affine equivalence is 302 [2], among which 10 s-boxes have nonlinearity 4, degree 3 and absolute autocorrelation value 8, which are the basic cryptographic properties that make these s-boxes useful in practice. We compute the TO values of each of these 10 s-boxes (denoted as F) and their extended affine equivalent ones (denoted as G). The results are shown in Table 1. From these tables it is evident that better TO values are obtained by considering the extended affine equivalent functions of these 10 s-boxes, hence we recommend to use such functions G from Table 1 that give better resistance against DPA, *i.e.,* lesser TO value. Following the Table 1, one should use some G such that $\mathsf{TO}(G) = 2.233$.

6 Practical Soundness of the Transparency Order

6.1 Attack Simulations

This section aims to confront the notion of (revised) transparency order with attack simulations. Essentially, our goal is to study to what extent the low transparency order of an s-box impacts the efficiency of a side channel attack against its processing.

s-box (F)	$\mathsf{TO}(F)$	β	s-box (G)	$\mathsf{TO}(G)$	β
084c2a1563efbd97	2.500	4	086ea29f45cb317d	2.266	2
084c2613a9db75ef	2.433	5	0e795f321ab6dc48	2.233	3
084c2a1563db79ef	2.400	3	08c4a29563dbf1e7	2.266	3
084c261da937b5ef	2.333	1	084c621de973f5ab	2.266	7
084c261da39b75ef	2.433	5	04519b826d7e3afc	2.233	4
084c261da3be9f57	2.533	7	0e97c58f24ab631d	2.233	5
084c2a1563ef7d9b	2.466	3	086e4c1725ab3f9d	2.266	2
084c261da3be5f97	2.533	6	092bd8a76c354fe1	2.333	0
084c261da3bef975	2.433	3	016794ebacf52d83	2.333	4
084c261da39b7e5f	2.433	7	04518a936f7c2bed	2.233	3

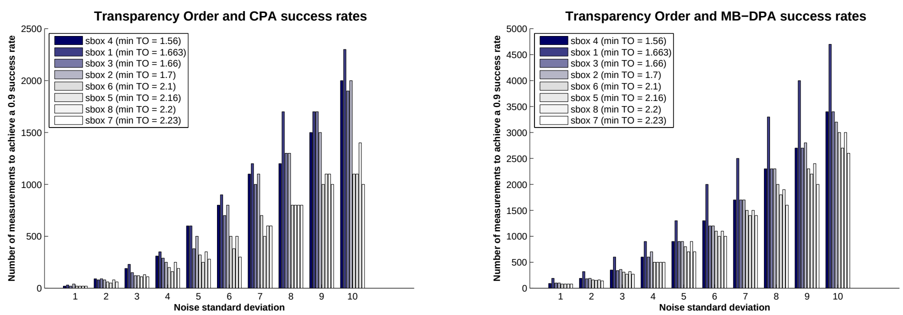
Table 1. All 4×4 s-box F up to affine equivalence with nonlinearity 4, degree 3 and absolute autocorrelation value 8, and their respective affine equivalent s-boxes that achieve the minimum $\mathsf{TO}(G)$; the β values are also given where the TO is attained.

We first performed `CPA` attack simulations against the 8 `PRINCE` s-boxes listed in Table 2 (Appendix A). We think that the latter ones are good targets for our study since their minimum transparency order are reasonably different and ranges from 1.56667 (for s-box 4) to 2.23333 (for s-box 7). In these first tests campaign, we choose to simulate the information leakage in the classical Hamming Distance model with Gaussian noise. Namely, the leakage $L(X \oplus \dot{K})$ related to the processing of the s-box output $F(X + \dot{K})$ equals $\mathsf{H}(F(x \oplus \dot{K}) \oplus \beta) + B$, where B is a random variable whose distribution is Gaussian with null mean and standard deviation σ. The value β corresponds to the initial state of the memory before the writing of $F(x \oplus \dot{K})$. According to the discussion in previous sections, we assumed that it can be chosen by the designer and, for each `PRINCE` s-box F, we selected it to minimize $\mathsf{TO}(F, \beta)$ (see Table 2)[6]. Each hypothesis K on $\dot{K}$ has been tested by estimating the correlation coefficient [7] $\rho(\mathsf{HW}(F(X \oplus K) \oplus \beta), L(X \oplus \dot{K}))$. It can be noticed that the initial content β of the register is assumed to be known by the attacker, which makes sense since it is part of the design parameters and therefore must be public according to Kerckhoff's rule. The number of leakage observations used to estimate the correlation is denoted by N. Attacks have been tested for different amounts of noise (namely for different standard deviations $\sigma \in [1..10]$). For each of them, we estimated the minimum number of observations N required for the attack to succeed with a probability at least equal to 0.9. As argued in [17, 25], this is a sound way to evaluate the efficiency of a side-channel attack. Results are reported in Figure 3(a).

[6] this should correspond to a maximum level of security.

[7] We recall that the correlation coefficient between two random variables U and V can be soundly estimated from respectively N observations $(u_i)_i$ and $(v_i)_i$ of U and V by $\rho(U, V) \simeq (N \sum_i u_i v_i - \sum_i u_i v_i)/(\sqrt{N \sum_i u_i^2 - (\sum_i u_i)^2}\sqrt{N \sum_i v_i^2 - (\sum_i v_i)^2})$.

Remark 6. Since leakages are generated with the same (Hamming distance) model used to compute the predictions, the use of the correlation coefficient as a statistical distinguisher is an optimal choice. In particular, we experimentally validated that replacing it by a mutual information (as proposed in [14]) does not improve the attacks success rates, nor change the conclusions we have drawn from our simulations with `CPA`.



(a) `CPA` (b) `MB-DPA`

Fig. 3. Minimum number of Messages (in y-axis) required to achieve a 90% success rate *versus* the noise standard deviation (in x-axis)

It may be checked in Figure 3(a) that the transparency order impacts the `CPA` attack efficiency in the Hamming distance model. This impact increases with the noise and, for $\sigma = 10$, the attack efficiency (*i.e.,* the number of traces) is almost multiplied by 2.5 if we compare s-boxes 1 and 7. One can also observe that $\mathsf{TO}_{\min}$ alone does not fully capture the resistance against `CPA` since sbox 1 seems to be always more resistant than sbox 4 whereas its $\mathsf{TO}_{\min}$ is slightly greater (1.663 *versus* 1.56).

In Section 5, we related the new notion of transparency to the multi-bit `DPA` attack introduced in [1]. Since the latter attack is not equivalent to a `CPA`, we ran a second attacks simulation campaign. The results are reported in Figure 3(b). As expected, they essentially confirm the results we had with the `CPA`: the lower $\mathsf{TO}_{\min}(F)$, the higher the resistance against the attacks.

In the second phase of our simulations, we wanted to investigate the robustness of the transparency order criterion against stochastic errors. In other words, we studied the impact of an erroneous modelling on the s-box `CPA` resistance, by performing `CPA` attack simulations against the fourth and the seventh `PRINCE` s-boxes[8] under the assumption that the information is not leaking in the Hamming distance model but in an erroneous version of it. Namely, for a fixed *stochastic*

[8] those s-boxes correspond to the two opposite extrema in terms of $\mathsf{TO}_{\min}(F)$.

error standard deviation σ_{er} chosen[9] in $\{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$, we simulated the leakage $L(X \oplus \dot{K})$ such that:

$$L(X \oplus \dot{K}) = \varphi(F(X \oplus \dot{K}) \oplus \beta) + B \ , \tag{16}$$

where φ is a function defined for every $y \in \mathbb{F}_2^4$ by $\varphi(y) = \mathrm{HW}(y) + \varepsilon$ with ε randomly generated according to a normal distribution with mean 0 and standard deviation σ_{er}. The variable B still refers to an independent Gaussian noise with 0 mean and standard deviation σ. For the processing of the predictions, we kept the Hamming weight model (the adversary is not assumed to know the erroneous leakage model). The results of our `CPA` attack simulations are reported in Figure 4 (bins in dark blue correspond to s-box 4 whereas those in light blue correspond to s-box 7, for each standard deviation σ – in x-axis – there is one bin for each stochastic error σ_{er} in $\{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$).

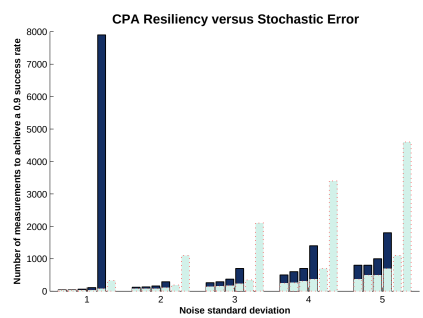


Fig. 4. `CPA` in presence of stochastic error - Minimum number of Messages (in y-axis) required to achieve a 90% success rate *versus* the noise standard deviation (in x-axis)

It may be checked that the fourth s-box, which has minimum $\mathsf{TO}_{\min}(F)$, stays more resistant than the seventh s-box for any stochastic error and the noise standard deviation. More interestingly, our simulations show that the difficulty of attacking s-box 4 increases more quickly with the stochastic error than for s-box 7. Actually, for a stochastic error greater than or equal to 0.8, a 90% success rate was achieved against s-box 4 only when the noise standard deviation was equal to 1. For greater noise standard deviations (and for $\sigma_{er} \geq 0.8$), this success rate was never achieved by `CPA` attacks with less than 500 000 traces.

[9] These standard deviations correspond to $j\%$ of the mean $\mathsf{H}(y)$ when y ranges uniformly over $\mathbb{F}_2^4$ and $j \in \{0, 10, 20, 30, 40, 50\}$.

6.2 Conclusion of the Practical Soundness of the Transparency Order

As shown by our simulations, the (minimum) transparency order is indeed related to the resistance of the s-box implementation against side channel attacks like the `CPA` or the multi-bit `DPA`. Choosing s-box with the minimum transparency order and using precharge value β for which the minimum is achieved seems therefore a good defence strategy. From this point of view, our simulations confirm our theoretical analysis. However, our simulations also show that a small minimum transparency order is not sufficient alone to achieve a satisfying resistance level against `CPA`: in the most favourable situation (Figure 3(a), no stochastic error and a great amount of noise), the number of needed observations to attack the s-box output is "only" multiplied by 2.5 when considering the two extreme cases of s-boxes 1 and 7. This is definitely not sufficient in practice where one usually expects that no attack succeeds with less than 1 million observations (or even more). As a conclusion, choosing s-boxes with small minimum transparency order is a sound strategy if it is combined with other classical countermeasures like *e.g.,* masking [7], shuffling [28] or threshold implementation [21]. Moreover, our analysis (*e.g.,* Table 1) suggests that among s-boxes with equal (and good) cryptographic properties, there may exist significant differences in terms of (minimum) transparency order.

7 Conclusion

In this paper we have critically analysed the definition of transparency order originally introduced in [24] almost a decade back. Even if several works have been published on this notion (*e.g.,* [5, 11, 18, 22]), we exhibited several inconsistencies in the definition as well as in the interpretation of the definition that went unnoticed for a long time. We have then conducted an in-depth analysis of the notion which led us to output a revised definition which answers all the issues identified in the previous version. Then, we critically investigated the practical soundness of the (revised) transparency order notion. Through several attacks simulation campaigns, we have shown that it is indeed related to the efficiency of side channel attacks like `CPA` and multi-bit `DPA` (which confirm the theoretical analyses in [24] and this paper). Actually, this result is also in line with the metrics studied by Whitnall and Oswald in [31], which share several similarities with our notion of transparency order[10]. Our simulations also showed that a small transparency order is not sufficient alone to ensure the practical security of the implementation and must therefore be combined with other classical countermeasures: combined with the latter ones it will lead to a security improvement.

As a final conclusion of this work, we think that the design of s-boxes with small transparency order is an interesting open avenue for further research. Also, it will be interesting to precisely study how the (minimum) transparency order

[10] Note that the metrics in [31] have been recently critically analysed in [27].

impacts the efficiency of attacks like the template attacks [9], the linear regression attacks [10] or the MIA [14], especially when it is combined with masking. Eventually, the (revised) notion of transparency order seems to share several similarities with the notion of *confusion coefficient* introduced in [12] and recently used in [23] to design `DPA` resistant s-boxes. A comparison analysis between the two approaches appears to be a promising subject for future research on this topic

References

1. R. Bévan and E. Knudsen. Ways to Enhance Power Analysis. In P. Lee and C. Lim, editors, *Information Security and Cryptology – ICISC 2002*, volume 2587 of *Lecture Notes in Computer Science*, pages 327–342. Springer, 2002.
2. B. Bilgin, S. Nikova, V. Nikov, V. Rijmen, and G. Stütz. Threshold implementations of all 3x3 and 4x4 s-boxes. In Prouff and Schaumont [26], pages 76–91.
3. J. Borghoff, A. Canteaut, T. Güneysu, E. B. Kavun, M. Knezevic, L. R. Knudsen, G. Leander, V. Nikov, C. Paar, C. Rechberger, P. Rombouts, S. S. Thomsen, and T. Yalçin. PRINCE - A low-latency block cipher for pervasive computing applications - extended abstract. In X. Wang and K. Sako, editors, *Advances in Cryptology - ASIACRYPT 2012 - 18th International Conference on the Theory and Application of Cryptology and Information Security, Beijing, China, December 2-6, 2012. Proceedings*, volume 7658 of *Lecture Notes in Computer Science*, pages 208–225. Springer, 2012.
4. E. Brier, C. Clavier, and F. Olivier. Correlation Power Analysis with a Leakage Model. In M. Joye and J.-J. Quisquater, editors, *Cryptographic Hardware and Embedded Systems – CHES 2004*, volume 3156 of *Lecture Notes in Computer Science*, pages 16–29. Springer, 2004.
5. C. Carlet. On Highly Nonlinear S-boxes and Their Inability to Thwart DPA Attacks. In S. Maitra, C. E. Veni Madhavan, and R. Venkatesan, editors, *Progress in Cryptology – INDOCRYPT 2005*, volume 3797 of *Lecture Notes in Computer Science*, pages 49–62. Springer, 2006.
6. C. Carlet. *Vectorial Boolean Functions for Cryptography*, chapter 9, pages 398–469. Boolean Models and Methods in Mathematics, Computer Science, and Engineering. Cambridge University Press, June 2010.
7. S. Chari, C. Jutla, J. Rao, and P. Rohatgi. A Cautionary Note Regarding Evaluation of AES Candidates on Smart-Cards. In *Second AES Candidate Conference – AES 2*, Mar. 1999.
8. S. Chari, C. Jutla, J. Rao, and P. Rohatgi. Towards Sound Approaches to Counteract Power-Analysis Attacks. In Wiener [32], pages 398–412.
9. S. Chari, J. Rao, and P. Rohatgi. Template Attacks. In B. Kaliski Jr., Ç. Koç, and C. Paar, editors, *Cryptographic Hardware and Embedded Systems – CHES 2002*, volume 2523 of *Lecture Notes in Computer Science*, pages 13–29. Springer, 2002.
10. J. Doget, E. Prouff, M. Rivain, and F.-X. Standaert. Univariate Side Channel Attacks and Leakage Modeling. *Journal of Cryptographic Engineering*, 1(2):123–144, 2011.
11. M. A. Evci and S. Kavut. DPA Resilience of Rotation-Symmetric S-boxes. In *IWSEC*, pages 146–157, 2014.
12. Y. Fei, Q. Luo, and A. A. Ding. A Statistical Model for DPA with Novel Algorithmic Confusion Analysis. In Prouff and Schaumont [26], pages 233–250.

13. FIPS PUB 197. *Advanced Encryption Standard*. National Institute of Standards and Technology, Nov. 2001.
14. B. Gierlichs, L. Batina, P. Tuyls, and B. Preneel. Mutual information analysis. In E. Oswald and P. Rohatgi, editors, *CHES*, volume 5154 of *Lecture Notes in Computer Science*, pages 426–442. Springer, 2008.
15. S. Guilley, P. Hoogvorst, and R. Pacalet. Differential Power Analysis Model and Some Results. In J.-J. Quisquater, P. Paradinas, Y. Deswarte, and A. E. Kalam, editors, *Smart Card Research and Advanced Applications VI – CARDIS 2004*, pages 127–142. Kluwer Academic Publishers, 2004.
16. P. Kocher, J. Jaffe, and B. Jun. Differential Power Analysis. In Wiener [32], pages 388–397.
17. S. Mangard. Hardware Countermeasures against DPA – A Statistical Analysis of Their Effectiveness. In T. Okamoto, editor, *Topics in Cryptology – CT-RSA 2004*, volume 2964 of *Lecture Notes in Computer Science*, pages 222–235. Springer, 2004.
18. B. Mazumdar, D. Mukhopadhyay, and I. Sengupta. Constrained search for a class of good bijective s-boxes with improved DPA resistivity. *IEEE Transactions on Information Forensics and Security*, 8(12):2154–2163, 2013.
19. W. Meier and D. Mukhopadhyay, editors. *Progress in Cryptology - INDOCRYPT 2014 - 15th International Conference on Cryptology in India, New Delhi, India, December 14-17, 2014, Proceedings*, volume 8885 of *Lecture Notes in Computer Science*. Springer, 2014.
20. T. Messerges. *Power Analysis Attacks and Countermeasures for Cryptographic Algorithms*. PhD thesis, University of Illinois, 2000.
21. S. Nikova, V. Rijmen, and M. Schläffer. Secure hardware implementation of non-linear functions in the presence of glitches. *J. Cryptology*, 24(2):292–321, 2011.
22. S. Picek, B. Ege, K. Papagiannopoulos, L. Batina, and D. Jakobovic. Optimality and beyond: The case of 4×4 s-boxes. In *2014 IEEE International Symposium on Hardware-Oriented Security and Trust, HOST 2014, Arlington, VA, USA, May 6-7, 2014*, pages 80–83. IEEE Computer Society, 2014.
23. S. Picek, K. Papagiannopoulos, B. Ege, L. Batina, and D. Jakobovic. Confused by confusion: Systematic evaluation of DPA resistance of various s-boxes. In Meier and Mukhopadhyay [19], pages 374–390.
24. E. Prouff. DPA attacks and S-Boxes. In H. Handschuh and H. Gilbert, editors, *Fast Software Encryption – FSE 2005*, volume 3557 of *Lecture Notes in Computer Science*, pages 424–442. Springer, 2005.
25. E. Prouff, M. Rivain, and R. Bévan. Statistical Analysis of Second Order Differential Power Analysis. *IEEE Transactions on Computers*, 58(6):799–811, 2009.
26. E. Prouff and P. Schaumont, editors. *Cryptographic Hardware and Embedded Systems - CHES 2012 - 14th International Workshop, Leuven, Belgium, September 9-12, 2012. Proceedings*, volume 7428 of *Lecture Notes in Computer Science*. Springer, 2012.
27. O. Reparaz, B. Gierlichs, and I. Verbauwhede. A note on the use of margins to compare distinguishers. In E. Prouff, editor, *Constructive Side-Channel Analysis and Secure Design - 5th International Workshop, COSADE 2014, Paris, France, April 13-15, 2014. Revised Selected Papers*, volume 8622 of *Lecture Notes in Computer Science*, pages 1–8. Springer, 2014.
28. M. Rivain, E. Prouff, and J. Doget. Higher-Order Masking and Shuffling for Software Implementations of Block Ciphers. In C. Clavier and K. Gaj, editors, *CHES*, volume 5747 of *Lecture Notes in Computer Science*, pages 171–188. Springer, 2009.

29. S. Sarkar, S. Maitra, and K. Chakraborty. Differential power analysis in hamming weight model: How to choose among (extended) affine equivalent s-boxes. In Meier and Mukhopadhyay [19], pages 360–373.
30. K. Tiri and I. Verbauwhede. A logic level design methodology for a secure DPA resistant ASIC or FPGA implementation. In *2004 Design, Automation and Test in Europe Conference and Exposition (DATE 2004), 16-20 February 2004, Paris, France*, pages 246–251. IEEE Computer Society, 2004.
31. C. Whitnall and E. Oswald. A Comprehensive Evaluation of Mutual Information Analysis Using a Fair Evaluation Framework. In P. Rogaway, editor, *CRYPTO*, volume 6841 of *Lecture Notes in Computer Science*, pages 316–334. Springer, 2011.
32. M. Wiener, editor. *Advances in Cryptology – CRYPTO '99*, volume 1666 of *Lecture Notes in Computer Science*. Springer, 1999.

A Analysis for the s-boxes in context of `PRINCE`

Eight 4×4 s-boxes are referred in [3]. In Table 2 we show the maximum and minimum values of $\mathsf{TO}(F, \beta)$ for each of the s-boxes.

s-box	$\beta_{\max}$ (as integer)	$\mathsf{TO}(F)$	$\beta_{\min}$ (as integer)	$\mathsf{TO}_{\min}(F)$
s-box-1	0	2.46667	1	1.63333
s-box-2	2	2.56666	1	1.7
s-box-3	2	2.53333	1	1.66667
s-box-4	4	2.46667	1	1.56667
s-box-5	4	2.53333	2	2.16667
s-box-6	0	2.46667	6	2.1
s-box-7	6	2.5	5	2.23333
s-box-8	2	2.66667	7	2.2

Table 2. Maximum (corresponding to $\beta_{\max}$) and minimum (corresponding to $\beta_{\min}$) values of $\mathsf{TO}(F, \beta)$ as β varies over $\mathbb{F}_2^4$ for the eight `PRINCE` s-boxes (available in Table 3 of Appendix A in the eprint version of [3]).

The following sections are given to help the review process but are not planned to be included in the final version.

B Detailed calculations

$\varDelta_{K,\dot{K}}(j,\beta)$ in terms of s-box parameters

$$\begin{aligned}\varDelta_{K,\dot{K}}(j,\beta) &= -\frac{1}{2^{n-1}}\sum_{x\in\mathbb{F}_2^n}(-1)^{F_j(x\oplus K)}\mathsf{H}\left(F(x\oplus\dot{K})\oplus\beta\right) \qquad (17)\\ &= -\frac{1}{2^{n-1}}\sum_{x\in\mathbb{F}_2^n}(-1)^{F_j(x\oplus K)}\sum_{i=1}^{m}(F_i(x\oplus\dot{K})\oplus\beta_i)\\ &= -\frac{1}{2^{n-1}}\sum_{x\in\mathbb{F}_2^n}(-1)^{F_j(x\oplus K)}\ \frac{1}{2}\left(m-\sum_{i=1}^{m}(-1)^{F_i(x\oplus\dot{K})\oplus\beta_i}\right)\\ &= -\frac{m}{2^n}\sum_{x\in\mathbb{F}_2^n}(-1)^{F_j(x\oplus K)}+\frac{1}{2^n}\sum_{i=1}^{m}(-1)^{\beta_i}\sum_{x\in\mathbb{F}_2^n}(-1)^{F_j(x\oplus K)\oplus F_i(x\oplus\dot{K})}\end{aligned}$$

Calculations related to $\varDelta_{K,\dot{K}}(j,\beta)$

$$\varDelta_{K,\dot{K}}(j,\beta)=Q_1+Q_2,\ \text{where} \qquad (18)$$

$$\begin{aligned}Q_1 &= \frac{1}{2^n}\sum_{x\in\{0,1\}^n}(-1)^{\beta_j}(-1)^{F_j(x\oplus K)\oplus F_j(x\oplus\dot{K})},\ \text{for } i=j\\ &= \frac{(-1)^{\beta_j}}{2^n}\mathcal{A}_{F_j}(K\oplus\dot{K}),\ \text{and}\end{aligned}$$

$$\begin{aligned}Q_2 &= \frac{1}{2^n}\sum_{i=1,i\neq j}^{m}(-1)^{\beta_i}\sum_{x\in\{0,1\}^n}(-1)^{F_j(x\oplus K)\oplus F_i(x\oplus\dot{K})},\ \text{for } i\neq j\\ &= \frac{1}{2^n}\sum_{i=1,i\neq j}^{m}(-1)^{\beta_i}\mathcal{C}_{F_j,F_i}(K\oplus\dot{K}).\end{aligned}$$

Proof of ${\tau'}_F^{\overline{\beta}}={\tau'}_F^{\beta}$

$$\begin{aligned}
{\tau'}^{\overline{\beta}}_F &= |m - 2\mathsf{H}(\overline{\beta})| \\
&\quad - \frac{1}{2^n(2^n-1)} \sum_{a \in {\mathbb{F}_2^n}^*} \left| \sum_{j=1}^{m} \left((-1)^{\overline{\beta}_j} \mathcal{A}_{F_j}(a) + \sum_{i=1, i\neq j}^{m} (-1)^{\overline{\beta}_i} \mathcal{C}_{F_i,F_j}(a) \right) \right| \\
&= |m - 2\mathsf{H}(\beta)| \\
&\quad - \frac{1}{2^n(2^n-1)} \sum_{a \in {\mathbb{F}_2^n}^*} \left| \sum_{j=1}^{m} \left(-(-1)^{\beta_j} \mathcal{A}_{F_j}(a) - \sum_{i=1, i\neq j}^{m} (-1)^{\beta_i} \mathcal{C}_{F_i,F_j}(a) \right) \right| \\
&= |m - 2\mathsf{H}(\beta)| \\
&\quad - \frac{1}{2^n(2^n-1)} \sum_{a \in {\mathbb{F}_2^n}^*} \left| \sum_{j=1}^{m} \left((-1)^{\beta_j} \mathcal{A}_{F_j}(a) + \sum_{i=1, i\neq j}^{m} (-1)^{\beta_i} \mathcal{C}_{F_i,F_j}(a) \right) \right| \\
&= {\tau'}^{\beta}_F. \qquad (19)
\end{aligned}$$

Proof of Theorem 2

Theorem 3. *Let F be an $n \times m$ balanced function. Then, for any affine permutation $A \in \mathbf{A}_n$ we have $\mathsf{TO}(F \circ A) = \mathsf{TO}(F)$.*

Proof. Suppose $F = (F_1, \ldots, F_n)$. Equation (14) implies:

$$\mathsf{TO}(F \circ A, \beta) = m - \frac{1}{2^n(2^n-1)} \sum_{a \in {\mathbb{F}_2^n}^*} \sum_{j=1}^{m} \left| \sum_{i=1}^{m} (-1)^{\beta_i \oplus \beta_j} \mathcal{C}_{F_i \circ A, F_j \circ A}(a) \right| \quad . \qquad (20)$$

Since A is an affine permutation over $\mathbb{F}_2^n$, we have

$$\mathcal{C}_{F_i \circ A, F_j \circ A}(a) = \sum_{x \in \mathbb{F}_2^n} (-1)^{F_i \circ A(x) \oplus F_j \circ A(x \oplus a)} = \sum_{x \in \mathbb{F}_2^n} (-1)^{F_i(x) \oplus F_j(x \oplus A(a) \oplus A(0))} \ ,$$

that is

$$\mathcal{C}_{F_i \circ A, F_j \circ A}(a) = \mathcal{C}_{F_i, F_j}(L(a)) \ , \qquad (21)$$

where L is the linear function defined by $L(a) = A(a) \oplus A(0)$.

Equations (20) and (21) together straightforwardly imply:

$$\mathsf{TO}(F \circ A, \beta) = m - \frac{1}{2^n(2^n-1)} \sum_{a \in {\mathbb{F}_2^n}^*} \sum_{j=1}^{m} \left| \sum_{i=1}^{m} (-1)^{\beta_i \oplus \beta_j} \mathcal{C}_{F_i, F_j}(a) \right|$$

i.e., $\mathsf{TO}(F \circ A, \beta) = \mathsf{TO}(F, \beta)$ which concludes the proof. $\square$

C A lower bound of $\mathsf{TO}(F)$ using Walsh spectrum only

We present a lower bound of $\mathsf{TO}(F)$ for a given $F = (F_1, \ldots, F_m)$. The following result will be used to derive the bound.

Lemma 1. *Suppose e, f, g, h are Boolean functions of n-variables. Then*

$$\sum_{a \in \mathbb{F}_2^n} \mathcal{C}_{e,f}(a)\mathcal{C}_{g,h}(a) = \frac{1}{2^n} \sum_{a \in \mathbb{F}_2^n} W_e(a)W_f(a)W_g(a)W_h(a).$$

Proof. Suppose $\mathbb{F}_2^n = \{a_0, \ldots, a_{2^n-1}\}$. It is known that

$$[\mathcal{C}_{e,f}(a_0), \ldots, \mathcal{C}_{e,f}(a_{2^n-1})]\mathcal{H}_n = [W_e(a_0)W_f(a_0), \ldots, W_e(a_{2^n-1})W_f(a_{2^n-1})]$$

$$[\mathcal{C}_{g,h}(a_0), \ldots, \mathcal{C}_{g,h}(a_{2^n-1})]\mathcal{H}_n = [W_g(a_0)W_h(a_0), \ldots, W_g(a_{2^n-1})W_h(a_{2^n-1})],$$

where $\mathcal{H}_n$ is the Hadamard matrix of order $2^n \times 2^n$. Take the product

$$\begin{aligned}
&[\mathcal{C}_{e,f}(a_0), \ldots, \mathcal{C}_{e,f}(a_{2^n-1})]\mathcal{H}_n\Big([\mathcal{C}_{g,h}(a_0), \ldots, \mathcal{C}_{g,h}(a_{2^n-1})]\mathcal{H}_n\Big)^T \\
&= [W_e(a_0)W_f(a_0), \ldots, W_e(a_{2^n-1})W_f(a_{2^n-1})] \begin{pmatrix} W_g(a_0)W_h(a_0 \\ \vdots \\ W_g(a_{2^n-1})W_h(a_{2^n-1}) \end{pmatrix}
\end{aligned}$$

Since, $\mathcal{H}_n\mathcal{H}_n^T = 2^n I_{2^n \times 2^n}$, where $I_{2^n \times 2^n}$ is the identity matrix of order $2^n \times 2^n$, then from the product, we have

$$\sum_{a \in \mathbb{F}_2^n} \mathcal{C}_{e,f}(a)\mathcal{C}_{g,h}(a) = \frac{1}{2^n} \sum_{a \in \mathbb{F}_2^n} W_e(a)W_f(a)W_g(a)W_h(a).$$

Theorem 4. *For $F = (F_1, \ldots, F_m) : \mathbb{F}_2^n \to \mathbb{F}_2^m$, the value of $\mathsf{TO}(F)$ has the following lower bound*

$$m - \tfrac{\sqrt{2^n-1}}{(2^{2n}-2^n)} \sum_{j=1}^{m} \left(\sum_{i=1}^{m} \sum_{a \in F_2^{n*}} W_{F_i}^2(a)W_{F_j}^2(a) + 2 \sum_{1 \le i < k \le m} \sum_{a \in F_2^{n*}} W_{F_i}(a)W_{F_j}^2(a)W_{F_k}(a) \right)^{\frac{1}{2}}.$$

Proof. It is clear that $\mathsf{TO}(F) \geq \mathsf{TO}(F, 0)$. So we calculate a lower bound of $\mathsf{TO}(F, 0)$. From (14) we get

$$\mathsf{TO}(F, 0) = m - \frac{1}{(2^{2n} - 2^n)} \sum_{j=1}^{m} \sum_{a \in F_2^{n*}} |\sum_{i=1}^{m} \mathcal{C}_{F_i,F_j}(a)| \tag{22}$$

Applying Cauchy-Schwarz inequality we get

$$\begin{aligned}
\sum_{a\in F_2^{n*}} |\sum_{i=1}^{m} \mathcal{C}_{F_i,F_j}(a)| &\leq \left((2^n-1)\sum_{a\in F_2^{n*}}\Big(\sum_{i=1}^{m}\mathcal{C}_{F_i,F_j}(a)\Big)^2\right)^{\frac{1}{2}} \\
&= \left((2^n-1)\sum_{a\in F_2^{n}}\left[\Big(\sum_{i=1}^{m}\mathcal{C}_{F_i,F_j}(a)\Big)^2-\Big(\sum_{i=1}^{m}\mathcal{C}_{F_i,F_j}(0)\Big)^2\right]\right)^{\frac{1}{2}} \\
&= \left((2^n-1)\sum_{a\in F_2^{n}}\Big(\sum_{i=1}^{m}\mathcal{C}_{F_i,F_j}(a)\Big)^2\right)^{\frac{1}{2}}
\end{aligned} \tag{23}$$

Note that

$$\begin{aligned}
\sum_{a\in F_2^{n}}\Big(\sum_{i=1}^{m}\mathcal{C}_{F_i,F_j}(a)\Big)^2 &= \sum_{a\in F_2^{n}}\sum_{i=1}^{m}\mathcal{C}^2_{F_i,F_j}(a)+2\sum_{a\in F_2^{n}}\sum_{1\leq i<k\leq m}\mathcal{C}_{F_i,F_j}(a)\mathcal{C}_{F_k,F_j}(a) \\
&= \sum_{i=1}^{m}\sum_{a\in F_2^{n}}\mathcal{C}^2_{F_i,F_j}(a)+2\sum_{1\leq i<k\leq m}\sum_{a\in F_2^{n}}\mathcal{C}_{F_i,F_j}(a)\mathcal{C}_{F_k,F_j}(a)
\end{aligned}$$

Then applying Lemma 1,

$$\sum_{a\in F_2^{n}}\Big(\sum_{i=1}^{m}\mathcal{C}_{F_i,F_j}(a)\Big)^2 = \sum_{i=1}^{m}\sum_{a\in F_2^{n}} W^2_{F_i}(a)W^2_{F_j}(a)+2\sum_{1\leq i<k\leq m}\sum_{a\in F_2^{n}} W_{F_i}(a)W^2_{F_j}(a)W_{F_k}(a).$$

Replacing this value of $\sum_{a\in F_2^{n}}\Big(\sum_{i=1}^{m}\mathcal{C}_{F_i,F_j}(a)\Big)^2$ in (23), an upper bound of $\sum_{a\in F_2^{n*}}|\sum_{i=1}^{m}\mathcal{C}_{F_i,F_j}(a)|$ is obtained. Then using this upper bound in (22), we get a lower bound of $\mathsf{TO}(F,0)$ as follows

$$m-\frac{\sqrt{2^n-1}}{(2^{2n}-2^n)}\sum_{j=1}^{m}\left(\sum_{i=1}^{m}\sum_{a\in F_2^{n}} W^2_{F_i}(a)W^2_{F_j}(a)+2\sum_{1\leq i<k\leq m}\sum_{a\in F_2^{n}} W_{F_i}(a)W^2_{F_j}(a)W_{F_k}(a)\right)^{\frac{1}{2}}.$$

Note that $\mathsf{TO}(F,\beta)$ assumes that all the coordinate functions are balanced, therefore the above bound can be written as

$$m-\frac{\sqrt{2^n-1}}{(2^{2n}-2^n)}\sum_{j=1}^{m}\left(\sum_{i=1}^{m}\sum_{a\in F_2^{n*}} W^2_{F_i}(a)W^2_{F_j}(a)+2\sum_{1\leq i<k\leq m}\sum_{a\in F_2^{n*}} W_{F_i}(a)W^2_{F_j}(a)W_{F_k}(a)\right)^{\frac{1}{2}}.$$

This serves as a lower bound of $\mathsf{TO}(F)$. $\square$

D Further exploring the Transparency Order: 1-st Order Distance

The balancedness of $F = (F_1, \cdots, F_m)$ implies that all the functions in the form $F_u \oplus F_v$, with u and v being distinct elements of $[1..m]$, are balanced. As originally observed in [24], this directly implies that the coefficient $\Delta_{K,\dot{K}}(v,\beta)$ equals $(-1)^{\beta_v}$ if $K = \dot{K}$. Based on this observation, it has been proposed in [29] to measure the resistance of an s-box against DPA attacks by computing the following Euclidean distance $\mathsf{d}^{(2)}_{K,\dot{K}}(\beta)$ between $((-1)^{\beta_v})_{v\in[1..m]}$ and the vector $(\Delta_{K,\dot{K}}(v,\beta)))_{v\in[1..m]}$ for each key candidate K:

$$\mathsf{d}^{(2)}_{K,\dot{K}}(\beta) = ||((-1)^{\beta_v})_{v\in[1..m]} - (\Delta_{K,\dot{K}}(v,\beta)))_{v\in[1..m]}||_2 \ ,$$

where $||\cdot||_2$ denotes the Euclidean norm.

The idea of [29] can obviously be extended to any norm. For instance, for the Manhattan norm $||\cdot||_1$ we can defined a new metric $\mathsf{d}^{(1)}_{K,\dot{K}}(\beta)$ such that:

$$\mathsf{d}^{(1)}_{K,\dot{K}}(\beta) = ||((-1)^{\beta_v})_{v\in[1..m]} - (\Delta_{K,\dot{K}}(v,\beta)))_{v\in[1..m]}||_1 \ .$$

Let us denote by $\mathsf{d}(F,\beta)$ the mean of the $\mathsf{d}^{(1)}_{K,\dot{K}}(\beta)$ for $K \neq \dot{K}$ (*i.e.,* $\mathsf{d}(\beta) = \frac{1}{2^n-1}\sum_{K\in\mathbb{F}_2^n, K\neq\dot{K}} \mathsf{d}^{(1)}_{K,\dot{K}}(\beta)$). We prove hereafter that for an s-box F, $\max_{\beta\in\mathbb{F}_2^n} \mathsf{d}(F,\beta)$ can serve as an upper bound for $\mathsf{TO}(F)$.

Proposition 4. *Let F be a balanced $n \times m$ s-box, then we have:*

$$\mathsf{TO}(F) \leqslant \max_{\beta\in\mathbb{F}_2^n} \mathsf{d}(F,\beta) \ .$$

Proof. By definition of $\mathsf{d}(\beta)$ we have

$$\mathsf{d}(F,\beta) = \frac{1}{(2^n-1)} \sum_{\substack{K\in\mathbb{F}_2^n\\ K\neq\dot{K}}} \sum_{v=1}^{m} \left|(-1)^{\beta_v} - \Delta_{K,\dot{K}}(v,\beta)\right| \quad i.e.,$$

$$\begin{aligned}\mathsf{d}(F,\beta) &\geq \frac{1}{2^n-1} \sum_{\substack{K\in\mathbb{F}_2^n\\ K\neq\dot{K}}} (m - \sum_{j=1}^{m} |\Delta_{K,\dot{K}}(j,\beta)|) \ , \\ &= \mathsf{TO}(F,\beta) \ \text{, by (13)} \ .\end{aligned}$$

Therefore, $\mathsf{TO}(F) = \max_{\beta\in\mathbb{F}_2^n} \mathsf{TO}(F,\beta) \leqslant \max_{\beta\in\mathbb{F}_2^n} \mathsf{d}(F,\beta)$. □

We went for a similar exercise as in Table 1 for $\max_{\beta\in\mathbb{F}_2^4} \mathsf{d}(\beta)$ instead of $\mathsf{TO}(F)$ (resp. $\mathsf{TO}(G)$). In each EA-equivalence class (*i.e.,* for each set of functions EA-equivalent to one of the 10 s-boxes F in Table 1), we obtained the minimum value 4.266 for $\min_{G\in\mathrm{EA}(F)} \max_{\beta\in\mathbb{F}_2^4} \mathsf{d}(\beta)$.