Journal of Cryptology

Methods and Instruments for Designing S-Boxes[1]

Réjane Forré
Institute for Communication Technology, Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland

Abstract. The design of S-boxes with minimal mutual information between input and output subvectors (considered as random variables) is investigated. First, the conditional entropy of the value of a boolean function conditioned on its random arguments is expressed as a function of the Walsh transform of the function. The *entropy profile* of a function is then defined; it allows the comparison of functions with regard to their (conditional) entropies. An algorithm to construct functions with good entropy profiles is then presented. It consists of a stepwise improvement of randomly chosen functions and uses the relation between the Walsh transform and the (conditional) entropies of a function. The statistical independency of boolean functions is investigated in the final section.

Key words. Product ciphers, S-boxes, Boolean functions, Walsh transform, Entropy, Statistical independency.

1. Introduction

The concept of *product cipher* was introduced by Feistel in [1]. The idea is "to combine two or more ciphers in such a way that the resulting system is stronger than either of the component systems alone." The Data Encryption Standard (DES) block cipher consists, for example, of 16 iterations of a relatively simple function f together with two fixed permutations. The so-called substitution boxes (S-boxes) are the only nonlinear components of the function f, and the strength of a block cipher with a DES-like structure relies heavily on the careful design of these S-boxes. Hereafter, we propose design instruments based on information theoretical considerations together with a heuristic and probabilistic optimization algorithm.

2. Boolean Functions with Large Entropies

2.1. *Conditional Entropies of a Boolean Function Expressed by Its Walsh Coefficients*

In the design of cipher systems, the end purpose is to minimize the mutual information $I(X; Y)$ between the plaintext X and the ciphertext Y (consdered as two random

[1] Date received: September 26, 1989. Date revised: February 26, 1990.

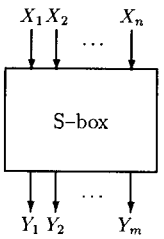

Fig. 1. A general substitution box (S-box).

variables). We propose to design S-boxes according to a similar principle. An S-box (Fig. 1) can be considered as an ordered set of m boolean functions, each of n boolean random variables $X_1, X_2, \ldots, X_n \in \{0, 1\}$:

$$f_1, f_2, \ldots, f_m: \mathrm{GF}(2)^n \to \{-1, 1\}$$

$$\mathbf{X} \mapsto Y_1, Y_2, \ldots, Y_m.$$

For computational reasons, we consider ± 1-valued instead of 0/1-valued boolean functions.

Since an S-box is a deterministic entity, it makes no sense to minimize the mutual information $I(\mathbf{X}; \mathbf{Y})$ between the random input vector $\mathbf{X} = (X_1 X_2 \cdots X_n)$ and the random output vector $\mathbf{Y} = (Y_1 Y_2 \cdots Y_m)$. But we can minimize mutual informations between input and output random *subvectors*. Let us consider, for example, the mutual information $I(X_i; Y_j)$ between the random input variable X_i and the random output variable Y_j. In the worst case, this mutual information could be one, which means that the random variable Y_j only depends on the input variable X_i. In the best case, $I(X_i; Y_j)$ should be zero, that is, X_i and Y_j should be statistically independent. But it is also important to minimize the mutual informations $I(\mathbf{X}'; Y_j)$ between input subvectors $\mathbf{X}'$ with $n' < n$ components and the output variable $Y_j (j = 1, 2, \ldots, m)$, as well as the mutual informations $I(Y_i; Y_j)$ between output variables Y_i and Y_j, $i, j \in \{1, 2, \ldots, m\}$. Minimizing these mutual informations is equivalent to maximizing the (conditional) entropies

$$H(Y_s), \tag{1}$$

$$H(Y_s | X_{i_1} \cdots X_{i_k}), \tag{2}$$

and

$$H(Y_s | Y_{j_1} \cdots Y_{j_l}) \tag{3}$$

for $s = 1, 2, \ldots, m$, $1 \le k \le n - 1$, $1 \le i_1 \cdots i_k \le n$, $1 \le l \le m - 1$, $j_1 \cdots j_l \ne s$, and $j_1 \cdots j_l \in \{1, 2, \ldots, m\}$. The entropies $H(Y_s)$, $s = 1, 2, \ldots, m$, obviously reach the maximum value 1 for balanced functions $f_s(X_1 X_2 \cdots X_n)$, that is, for functions mapping half the $\mathbf{X} \in \mathrm{GF}(2)^n$ onto 1 and the other half onto -1. The second expression to be maximized can be written as

$$H(Y_s | X_{i_1} \cdots X_{i_k}) = -\sum_{\substack{y_s \\ x_{i_1} \cdots x_{i_k}}} P(y_s, x_{i_1}, \ldots, x_{i_k}) \log_2 P(y_s | x_{i_1}, \ldots, x_{i_k}), \tag{4}$$

which can be simplified if we assume that all 2^k input subvectors $(x_{i_1}, \ldots, x_{i_k})$ are

equally probable, yielding

$$H(Y_s|X_{i_1}\cdots X_{i_k}) = -\sum_{\substack{y_s \\ x_{i_1}\cdots x_{i_k}}} 2^{-k}P(y_s|x_{i_1},\ldots,x_{i_k})\log_2 P(y_s|x_{i_1},\ldots,x_{i_k}) \quad (5)$$

$$= 2^{-k}\sum_{x_{i_1}\cdots x_{i_k}} h[P(Y_s = 1|x_{i_1},\ldots,x_{i_k})] \quad (6)$$

$$= 2^{-k}\sum_{x_{i_1}\cdots x_{i_k}} h[2^{k-n}\cdot \#\{\mathbf{X}: \mathbf{X}' = \mathbf{x}', f_s(\mathbf{X}) = 1\}], \quad (7)$$

where $h(p) = -p\log_2 p - (1-p)\log_2(1-p)$ denotes the binary entropy function, $\#\{\cdot\}$ denotes the cardinality of the enclosed set, and $\mathbf{X}' = (X_{i_1}X_{i_2}\cdots X_{i_k})$, $\mathbf{x}' = (x_{i_1}x_{i_2}\cdots x_{i_k})$. First, we consider the case $k = 1$:

$$H(Y_s|X_i) = \tfrac{1}{2}\cdot[h(2^{1-n}\cdot \#\{\mathbf{X}\in \mathrm{GF}(2)^n : X_i = 0, f_s(\mathbf{X}) = 1\}) + h(2^{1-n}\cdot \#\{\mathbf{X}\in \mathrm{GF}(2)^n : X_i = 1, f_s(\mathbf{X}) = 1\})]. \quad (8)$$

Introducing the notation

$$N_c^i(y) = \#\{\mathbf{X}\in \mathrm{GF}(2)^n : X_i = c, f_s(\mathbf{X}) = y\}, \qquad c\in\{0,1\}, \quad y\in\{1,-1\}, \quad (9)$$

equation (8) becomes

$$H(Y_s|X_i) = \tfrac{1}{2}[h(2^{1-n}N_0^i(1)) + h(2^{1-n}N_1^i(1))]. \quad (10)$$

The Walsh transform $F_s(\mathbf{w})$ of a boolean function f_s: $\mathrm{GF}(2)^n \to \{1,-1\}$ is defined by

$$F_s(\mathbf{w}) = \sum_{x\in \mathrm{GF}(2)^n} f_s(\mathbf{w})(-1)^{\mathbf{x}\cdot\mathbf{w}}, \quad (11)$$

with $\mathbf{x}\cdot\mathbf{w} = x_1w_1 \oplus x_2w_2 \oplus \cdots \oplus x_nw_n$. In particular

$$F_s(\mathbf{0}) = \sum_{x\in \mathrm{GF}(2)^n} f_s(\mathbf{x}) \quad (12)$$

$$= N_0^i(1) + N_1^i(1) - N_0^i(-1) - N_1^i(-1), \qquad i = 1,\ldots,n, \quad (13)$$

and, if we denote by $\mathbf{c}_i$ the vector with a one at the ith position and zeros at all other positions,

$$F_s(\mathbf{c}_i) = \sum_{x\in \mathrm{GF}(2)^n} f_s(\mathbf{x})(-1)^{x_i} \quad (14)$$

$$= N_0^i(1) - N_1^i(1) - N_0^i(-1) + N_1^i(-1). \quad (15)$$

Adding (resp. subtracting) (13) and (15) yields

$$N_0^i(1) - N_0^i(-1) = \tfrac{1}{2}[F_s(\mathbf{0}) + F_s(\mathbf{c}_i)] \quad (16)$$

and

$$N_1^i(1) - N_1^i(-1) = \tfrac{1}{2}[F_s(\mathbf{0}) - F_s(\mathbf{c}_i)]. \quad (17)$$

Using the fact that $N_0^i(1) + N_0^i(-1) = N_1^i(1) + N_1^i(-1) = 2^{n-1}$ we get

$$N_0^i(1) = 2^{n-2} + \tfrac{1}{4}[F_s(\mathbf{0}) + F_s(\mathbf{c}_i)], \quad (18)$$

$$N_0^i(-1) = 2^{n-2} - \tfrac{1}{4}[F_s(\mathbf{0}) + F_s(\mathbf{c}_i)], \quad (19)$$

$$N_1^i(1) = 2^{n-2} + \tfrac{1}{4}[F_s(\mathbf{0}) - F_s(\mathbf{c}_i)], \quad (20)$$

$$N_1^i(-1) = 2^{n-2} - \tfrac{1}{4}[F_s(\mathbf{0}) - F_s(\mathbf{c}_i)]. \quad (21)$$

Equation (10) becomes

$$H(Y_s|X_i) = \frac{1}{2}\left[h\left(\frac{1}{2} + \frac{F_s(\mathbf{0}) + F_s(\mathbf{c}_i)}{2^{n+1}}\right) + h\left(\frac{1}{2} + \frac{F_s(\mathbf{0}) - F_s(\mathbf{c}_i)}{2^{n+1}}\right)\right], \tag{22}$$

which expresses the conditional entropy of Y_s given X_i as a function of the Walsh coefficients $F_s(\mathbf{0})$ and $F_s(\mathbf{c}_i)$. We see that the conditional entropies of Y_s given one single input bit only depend on the values $F_s(\mathbf{w})$ for w's with Hamming weights ≤ 1.

The conditional entropy of Y_s given two input bits X_i and X_j can be computed in a similar way:

$$\begin{aligned} H(Y_s|X_iX_j) = 2^{-2}\cdot\{&h[\tfrac{1}{2} + 2^{-n-1}\cdot(F_s(\mathbf{0}) + F_s(\mathbf{c}_i) + F_s(\mathbf{c}_j) + F_s(\mathbf{c}_{ij}))] \\ &+ h[\tfrac{1}{2} + 2^{-n-1}\cdot(F_s(\mathbf{0}) + F_s(\mathbf{c}_i) - F_s(\mathbf{c}_j) - F_s(\mathbf{c}_{ij}))] \\ &+ h[\tfrac{1}{2} + 2^{-n-1}\cdot(F_s(\mathbf{0}) - F_s(\mathbf{c}_i) + F_s(\mathbf{c}_j) - F_s(\mathbf{c}_{ij}))] \\ &+ h[\tfrac{1}{2} + 2^{-n-1}\cdot(F_s(\mathbf{0}) - F_s(\mathbf{c}_i) - F_s(\mathbf{c}_j) + F_s(\mathbf{c}_{ij}))]\}, \end{aligned} \tag{23}$$

where $\mathbf{c}_{ij}$ denotes the vector with Hamming weight 2 and ones at the ith and jth positions. We see that the conditional entropy of Y_s given two input bits only depends on Walsh coefficients $F_s(\mathbf{w})$ with $H(\mathbf{w}) \leq 2$. We give the general formula without proof:

$$H(Y_s|X_{i_1}, X_{i_2}, \ldots, X_{i_k}) = 2^{-k}\sum_{x' \in \mathrm{GF}(2)^k} h\left(\frac{1}{2} + \frac{1}{2^{n+1}}\sum_{w' \in \mathrm{GF}(2)^k} F_s(O_n(\mathbf{w}'))(-1)^{\mathbf{w}'\cdot\mathbf{x}'}\right), \tag{24}$$

where $O_n(\mathbf{w}')$ denotes the n-dimensional vector $\mathbf{w}$ obtained by completing the k-dimensional subvector $\mathbf{w}'$ with zeros. For example, if $n = 6$, $k = 3$, $i_1 = 2$, $i_2 = 3$, $i_3 = 5$, $\mathbf{w}' = (1, 1, 1)$, we get $O_6(\mathbf{w}') = (0, 1, 1, 0, 1, 0)$.

Equation (24) confirms what we already observed for $k = 1$ and $k = 2$: the conditional entropy of $f_s(\mathbf{x})$ given k input bits $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ only depends on the Walsh transform $F_s(\mathbf{w})$ calculated for vectors $\mathbf{w}$ with Hamming weights $H(\mathbf{w}) \leq k$ $(0 \leq k \leq n-1)$. The special case where $H(Y|X_{i_1}\cdots X_{i_k})$ reaches its maximum value 1 for all choices of $i_1, i_2, \ldots, i_k \in [1\cdots n]$, that is, where the output Y_s is statistically independent of every set of k input bits $X_{i_1}\cdots X_{i_k}$, is called *correlation immunity of order k* [7] and its spectral characterization was studied in [10].

Figure 2 shows $H(Y_s|X_i)$ versus $F_s(\mathbf{c}_i)$ (for a particular $i \in \{1, 2, \ldots, n\}$) for a function f_s of $n = 6$ bits. The 33 curves correspond to the 33 possible values of $|F_s(\mathbf{0})|$ $(0 \leq |F_s(\mathbf{0})| \leq 2^6, F_s(\mathbf{0})$ even). The presence of two empty areas below the line $H(Y_s|X_i) = 0.5$ is due to the fact that, for a given $F_s(\mathbf{0})$, not all values of $F_s(\mathbf{c}_i)$ are possible. The restrictions arise from the conditions $N_0^i(1), N_0^i(-1), N_1^i(1), N_1^i(-1) \geq 0$. Equations (18)–(21) thus imply

$$|F_s(\mathbf{0}) + F_s(\mathbf{c}_i)| \leq 2^n \tag{25}$$

and

$$|F_s(\mathbf{0}) - F_s(\mathbf{c}_i)| \leq 2^n, \qquad \forall \mathbf{c}_i \in \mathrm{GF}(2)^n: H(\mathbf{c}_i) = 1. \tag{26}$$

Moreover, we know that only even values can appear in the Walsh transform of a boolean function [8, p. 168]. The curves of Fig. 2 show that, for a given $F_s(\mathbf{0})$, the conditional entropy $H(Y|X_i)$ monotonically increases when $|F_s(\mathbf{c}_i)|$ decreases. The

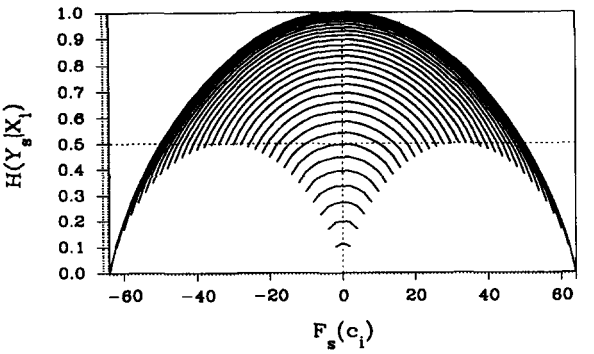

Fig. 2. The conditional entropy $H(Y_s|X_i)$ versus $F_s(\mathbf{c}_i)$ ($\forall i \in \{1, 2, \ldots, n\}$) for a function f_s of $n = 6$ bits. The uppermost curve corresponds to the case $F_s(\mathbf{0}) = 0$.

maximum $H(Y|X_i) = 1$ is only reached when $F_s(\mathbf{0}) = 0$ (i.e., f_s is balanced, see (12)) and $F_s(\mathbf{c}_i) = 0$, as pointed out in [10].

Figure 3 shows three collections of curves $H(Y_s|X_i, X_j)$ versus $F_s(\mathbf{c}_{ij})$, corresponding to the three cases $|F_s(\mathbf{c}_i)| = |F_s(\mathbf{c}_j)| = 0$, $|F_s(\mathbf{c}_i)| = |F_s(\mathbf{c}_j)| = 12$, and $|F_s(\mathbf{c}_i)| = |F_s(\mathbf{c}_j)| = 24$. Each curve corresponds to a given value of $|F_s(\mathbf{0})|$: zero for the uppermost curve, larger values for lower curves. As for $F_s(\mathbf{c}_i)$, there are restrictions on the

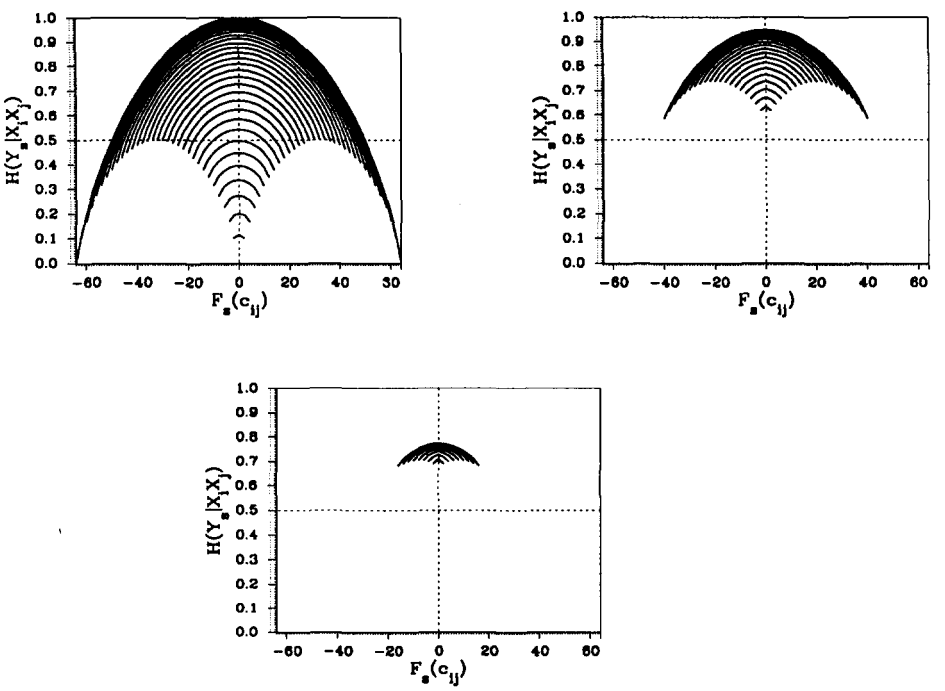

Fig. 3. The conditional entropy $H(Y_s|X_i, X_j)$ versus $F_s(\mathbf{c}_{ij})$ for a function f_s of $n = 6$ bits. The upper left graph corresponds to the case $|F_s(\mathbf{c}_i)| = |F_s(\mathbf{c}_j)| = 0$, the upper right graph corresponds to the case $|F_s(\mathbf{c}_i)| = |F_s(\mathbf{c}_j)| = 12$, and the lower graph corresponds to the case $|F_s(\mathbf{c}_i)| = |F_s(\mathbf{c}_j)| = 24$.

possible values of $F_s(\mathbf{c}_{ij})$ that can be derived in a similar way. These restrictions are given by

$$|F_s(\mathbf{0}) + F_s(\mathbf{c}_i) + F_s(\mathbf{c}_j) + F_s(\mathbf{c}_{ij})| \leq 2^n, \tag{27}$$

$$|F_s(\mathbf{0}) - F_s(\mathbf{c}_i) + F_s(\mathbf{c}_j) - F_s(\mathbf{c}_{ij})| \leq 2^n, \tag{28}$$

$$|F_s(\mathbf{0}) + F_s(\mathbf{c}_i) - F_s(\mathbf{c}_j) - F_s(\mathbf{c}_{ij})| \leq 2^n, \tag{29}$$

$$|F_s(\mathbf{0}) - F_s(\mathbf{c}_i) - F_s(\mathbf{c}_j) + F_s(\mathbf{c}_{ij})| \leq 2^n. \tag{30}$$

The curves of Fig. 2 show that, for given values of $F_s(\mathbf{0})$, $F_s(\mathbf{c}_i)$, and $F_s(\mathbf{c}_j)$, $H(Y|X_iX_j)$ monotonically increases when $F_s(\mathbf{c}_{ij})$ decreases. To attain the maximum $H(Y|X_iX_j) = 1$, $F_s(\mathbf{0})$, $F_s(\mathbf{c}_i)$, $F_s(\mathbf{c}_j)$, and $F_s(\mathbf{c}_{ij})$ must all be zero. If this requirement is judged too strong, we know from the curves of Figs. 2 and 3 that small values of $F_s(\mathbf{0})$, $F_s(\mathbf{c}_i)$, $F_s(\mathbf{c}_j)$, and $F_s(\mathbf{c}_{ij})$ also yield relatively high values of $H(Y|X_iX_j)$. This observation is important because it will allow us to construct "good" functions (i.e., functions with high values of $H(Y|X_i)$, $H(Y|X_iX_j)$, ...) without restricting us to the small class of correlation-immune functions of order 1, 2, ...

It should be clear that, even if we only treated the minimization of $H(Y|X_i)$ and $H(Y|X_iX_j)$, the results obtained can be extended to $H(Y|X_{i_1}X_{i_2}\cdots X_{i_k})$. The formal proof can be made by computing the partial derivatives of (24) with respect to $F_s(\mathbf{0})$, $F_s(\mathbf{c}_i)$, $F_s(\mathbf{c}_{ij})\cdots F_s(\mathbf{c}_{i_1\cdots i_k})$, but it is laborious and does not give any new interesting insight into the problem.

2.2. *An Algorithm To Construct Bollean Functions with Large Entropies*

We propose a probabilistic algorithm to construct boolean functions having good information theoretical properties, i.e., large (conditional) entropies. As a reference for "good" functions, we take the $8 \cdot 4 = 32$ boolean functions of 6 bits of the DES S-boxes. In order to make comparisons between different functions, we introduce the *entropy profile* of a function. It is a graph (see Fig. 4) showing, on the vertical axis, the values of $H(Y_s)$, $H(Y_s|X_i)$, $H(Y_s|X_iX_j)$, ..., $H(Y_s|X_{i_1}\cdots X_{i_{n-1}})$, where each block of values $H(Y_s|X_{i_1}\cdots X_{i_k})$ is sorted in descending order. In Fig. 4 the blocks are separated by dotted lines.

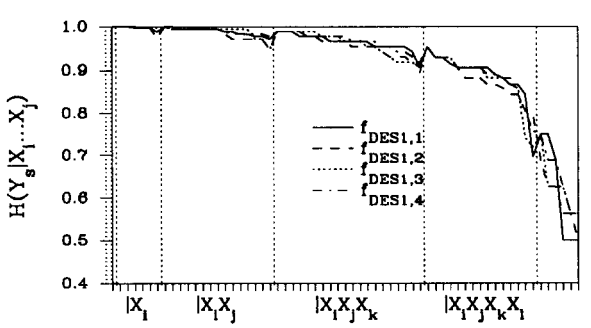

Fig. 4. Entropy profiles of the four functions of the DES S-box No. 1.

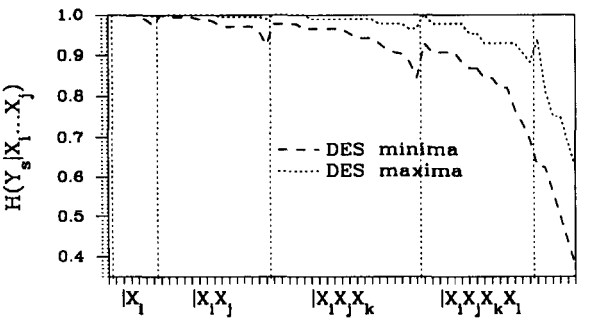

Fig. 5. Minimal and maximal entropies of the 32 functions of the eight DES S-boxes.

The graph of Fig. 5 shows the minima and the maxima of the entropies of the DES functions. Notice that the curve of minima, for example, is not characteristic of *one* function among the 32 DES functions, but each of its points is the lowermost point of all 32 entropy profile curves. The same is true for the curve of maxima on Fig. 5. Our design purpose will be to get functions whose entropy profiles lie as high as possible, without having any point below the DES minima curve. When constructing balanced boolean functions at random, we see that this criterion has a negligible probability of being fulfilled.

The method used in our algorithm is one of successive refinement of a randomly chosen start function f_0, until some given criteria are fulfilled. Its steps are briefly described hereafter.

1. Set $i = 0$. Choose a function f_0 of n bits completely at random (or randomly among a given class of functions, i.e., the class of bent functions [6], [5]).
2. Randomly flip one bit in the truth table of f_i to get a new function $g(\mathbf{x}) = f_i(\mathbf{x}) - 2 \cdot \delta(\mathbf{x} \oplus \mathbf{c}) f_i(\mathbf{x})$. Compute the Walsh spectra F_i and G of respectively f_i and g.
3. Compute the value $a(f_i, g)$, which denotes the quantitative improvement of g with respect to f_i. Roughly speaking, $a(f_i, g)$ grows when Walsh coefficients $G(\mathbf{w})$ are smaller in amplitude than the corresponding Walsh coefficients $F_i(\mathbf{w})$, for $\mathbf{w}$'s with Hamming weights 1, 2, ...
4. The modified function g replaces f_i with a probability p depending on the quantitative improvement $a(f_i, g)$:

$$\left.\begin{aligned} i &:= i + 1 \\ f_i(\mathbf{x}) &:= g(\mathbf{x}), \qquad \forall \mathbf{x} \in \mathrm{GF}(2)^n \\ F_i(\mathbf{w}) &:= G(\mathbf{w}), \qquad \forall \mathbf{x} \in \mathrm{GF}(2)^n \end{aligned}\right\} \text{Execute these steps with probability } p.$$

   If g was kept go to step 5, else go back to step 2.
5. Check whether the new f_i fulfills the following requirement:

$$\sum_{\substack{w \in \mathrm{GF}(2)^n: \\ H(w) = k}} |F_i(\mathbf{w})| \leq \max_i \sum_{\substack{w \in \mathrm{GF}(2)^n: \\ H(w) = k}} |F_{\mathrm{DES}_i}(\mathbf{w})| \tag{31}$$

for $k = 1, 2, \ldots, l, i = 1, \ldots, 32$. It is judicious to test whether (31) holds, because

it does not require as many computations as the determination of the whole entropy profile, and it allows us to avoid further testing of "bad" functions. If f_i passes the test go to step 6, else go back to step 2.

6. Flip the necessary number of bits in the truth table of f_i in such a way that f_i gets balanced and still keeps its large conditional entropies. Check whether the function obtained entirely lies above the DES minimal curve (of Fig. 5) and whether the distance of f_i to affine functions is large enough (see Section 2.3). If not, f_i is rejected and the algorithm failed.

This algorithm is quite heuristic: there is much freedom in defining the formula for the quantitative improvement of $a(f_i, g)$ in step 3, as well as for choosing the probability p as a function of a in step 4. In steps 4 and 5 we are free to decide up to which Hamming weight of w Walsh coefficients are to be taken into account. In step 6 we must choose a criterion to select which bits are to be flipped. We give hereafter a combination of parameters which yielded relatively good results, but there might be quite a few of other combinations leading to equivalent or even to better performances. In step 6 a lower limit for the distance $\delta(f_i)$ to affine functions must be set. For the S-boxes of the DES, the distances to affine functions lie between 14 and 22. For numbers $n \neq 6$ of arguments, the minimal allowed $\delta(f_i)$ could be typically chosen as some value close to 2^{n-2}. But this choice, again, is rather arbitrary and might be adapted according to the specific requirements of the designer. The essential characteristics of the algorithm are

- its probabilistic nature (allowing one given start function f_0 to lead to different end functions),
- the quantitative determination of the improvement of a function compared with another one looking at their respective Walsh spectra, and
- the (computationally) economic way of prechecking the "goodness" of the function in step 5 using sums over absolute values of its Walsh coefficients.

We experimented with this algorithm for functions of $n = 6$ bits, in order to compare the obtained functions with the DES functions. For different numbers of arguments, it would be necessary to modify step 5 of the algorithm, setting some well-chosen fixed lower limits for the entropy values (minimum entropy profile). One possibility of determining such a threshold entropy profile would be to use the above algorithm with the following modifications:

1. Execute steps 2–4 many times, so that the optimized function has a good chance of having a nice entropy profile.
2. Make the function balanced (as in step 6).
3. Keep the function only if it does not exhibit too deep "entropy canyons." A way of doing this is to check whether every point of the entropy profile does lie above a threshold of (for instance) 0.75 times the ordinate of the previous point.

A set of functions with possibly good entropy profiles is determined by means of the above procedure. The threshold entropy profile could then be taken as an average of the entropy profiles of the set, as the best entropy profile of the set, or

even as the curve consisting of the uppermost points of all the entropy profiles of the set. Only experiment is able to decide which one of these three choices is most adequate.

We computed the quantitative improvement $a(f_i, g)$ of g compared with f_i according to

$$a(f_i, g) = \tfrac{2}{3} \cdot \tfrac{1}{4} \cdot \left[n^{-1} \left(\sum_{\substack{w \in GF(2)^n: \\ H(w)=1}} |F_i(\mathbf{w})| - |G(\mathbf{w})| \right) + 2 \right] + \tfrac{1}{3} \cdot \tfrac{1}{4} \cdot \left[\binom{n}{2}^{-1} \left(\sum_{\substack{w \in GF(2)^n: \\ H(w)=2}} |F_i(\mathbf{w})| - |G(\mathbf{w})| \right) + 2 \right], \tag{32}$$

where $F_i(\mathbf{w})$ and $G(\mathbf{w})$ denote respectively the Walsh transforms of f_i and g. The multiplicative and additive constants in (32) were chosen in such a way that $a(f_i, g)$ takes values between 0 and 1, and that the first term has twice the weight of the second one. More importance is put on the first term because, as can be seen from Fig. 3, the best entropy values can only be reached if $|F_s(\mathbf{w})|$ is small (or vanishes) for $\mathbf{w}$'s with Hamming weight 1. If some values of $|F_s(\mathbf{w})|$ are large for $\mathbf{w}$'s with Hamming weight 1, small values of $|F_s(\mathbf{w})|$ at positions $\mathbf{w}$ with Hamming weight 2 do not permit us to obtain very good values of $H(Y_s | X_i, X_j)$. Of course, this does not justify the double weight of the first term in (32). To assign exact weights in (32), a sensitivity analysis of the conditional entropy $H(Y_s | X_{i_1} \cdots X_{i_k})$ with respect to the absolute values of $F(\mathbf{w})$ for $\mathbf{w}$'s with Hamming weights 1 and 2 should be performed. However, such an analysis is rather complex and its impact on the efficiency of the algorithm is difficult to foresee. Thus, we content ourselves with an arbitrary choice of parameters as in (32). Notice that the name of "improvement" is not particularly judicious for designating $a(f_i, g)$, since even a function g with poorer properties than f_i will be characterized by a positive (or zero) $a(f_i, g)$.

The probability p of acceptance in step 4 was taken as

$$p = \begin{cases} a^2(f_i, g) & \text{if } a(f_i, g) \leq 0.75, \\ \sqrt{a(f_i, g)} & \text{if } a(f_i, g) > 0.75. \end{cases} \tag{33}$$

Tests showed that too many "bad" modifications were accepted in step 4 if p was just taken equal to $a(f_i, g)$. The efficiency of the algorithm turned out to be better if, as in (33), the probability p of acceptance remains very small for small values of $a(f_i, g)$ (i.e., $a(f_i, g) \leq 0.75$) and gets large for large values of $a(f_i, g)$. In other words, the choice of formula (33) is based on experiments.

In step 5, sums over $\mathbf{w}$'s with Hamming weights 1 up to n were compared with the corresponding maximal sums of DES functions. To balance a function in step 6, we compared the functions $e_c(\mathbf{x})$ obtained by flipping the output of f_i at each possible position $\mathbf{c}$:

$$e_c(\mathbf{x}) := f_i(\mathbf{x}) - 2 \cdot \delta(\mathbf{x} \oplus \mathbf{c}) f_i(\mathbf{x}), \qquad \mathbf{c} \in GF(2)^n. \tag{34}$$

We then computed

$$T(\mathbf{c}) = \sum_{\substack{w \in GF(2)^n: \\ H(w)=1}} |E_c(\mathbf{w})| + \tfrac{2}{3} \sum_{\substack{w \in GF(2)^n: \\ H(w)=2}} |E_c(\mathbf{w})| + \tfrac{1}{5} \sum_{\substack{w \in GF(2)^n: \\ H(w)=3}} |E_c(\mathbf{w})| \tag{35}$$

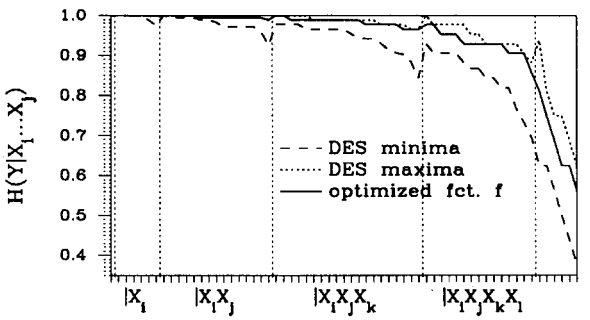

Fig. 6. Entropy profile of a function f determined by our algorithm compared with the minimal and maximal entropies of the 32 DES functions. The function f can be denoted by the hexadecimal number 936C9C96A699526B, which contains the $2^6 = 64$ successive outcomes of f.

and we flipped $f_i(\mathbf{c})$ at positions $\mathbf{c}$ with smallest values of $T(\mathbf{c})$. The necessary number of flips is given by $|F_i(\mathbf{0})|/2$. Note that the multiplicative constants in (35) (1, $\frac{2}{3}$, and $\frac{1}{5}$) were chosen for the same reasons that we put forward for (32): the smallness of $|F_s(\mathbf{w})|$ for $\mathbf{w}$'s with small Hamming weights seems to have more impact on the entropy profile of a function than the smallness of $|F_s(\mathbf{w})|$ for $\mathbf{w}$'s with large Hamming weights. The choice of these particular values (1, $\frac{2}{3}$, and $\frac{1}{5}$) in (35) was motivated by several trials of the algorithm with various multiplicative constants.

We tried out the algorithm with totally random functions as start functions as well as with randomly selected bent functions. The proportion of success was slightly better with bent functions (33.25% "good" functions) than with random functions (30.12% "good" functions). This might be explained by the fact that bent functions have flat Walsh spectra ($|F(\mathbf{w})| = 2^{n/2}$, $\forall \mathbf{w} \in \mathrm{GF}(2)^n$), and thus do not contain extreme large values of $|F(\mathbf{w})|$, that could be difficult to "flatten." But, as mentioned before, these performances strongly depend on the choice of the parameters and formulas in the algorithm and are therefore subject to improvement. What is important in our opinion is that the algorithm described is able to find functions with good entropy profiles much more efficiently than random search. Indeed, among 17,950 randomly selected balanced boolean functions, only 21 (i.e., 0.117%) were found to have entropy profiles without any point below the curve of DES minima. Figure 6 shows the entropy profile of a function that was found by the above algorithm, together with the minimal and maximal entropy curves of the DES functions.

2.3. *Comparison with Previous Work*

This section aims at describing the essential differences between previous characterizations of boolean functions (completeness, avalanche effect, strict avalanche criterion, nonlinearity) and the conditional entropies.

In [3] Kam and Davida introduce the idea of completeness: an $n \times n$ S-box (S: $\mathrm{GF}(2)^n \to \mathrm{GF}(2)^n$) is complete if, for every $i, j \in \{1, 2, \ldots, n\}$, there exists an input vector $\mathbf{x}$ such that $S(\mathbf{x})$ differs from $S(\mathbf{x} \oplus \mathbf{c}_i)$ at least in the jth bit. Obviously, the

completeness of an S-box simply requires the completeness of each of the boolean functions involved, which is defined as follows.

Definition 1. A boolean function f: GF(2)$^n \to \{1, -1\}$ is complete if and only if, for every $i \in \{1, 2, \ldots, n\}$, there exists an input vector $\mathbf{x}$ such that $f(\mathbf{x}) \neq f(\mathbf{x} \oplus \mathbf{c}_i)$, where $\mathbf{c}_i$ denotes the vector of GF(2)^n with a one at the ith position and zeros elsewhere.

In terms of entropy, the completeness of a boolean function f is expressed as

$$H(f(\mathbf{x} \oplus \mathbf{c}_i) | f(\mathbf{x})) \geq h(2^{1-n}), \qquad \forall i \in \{1, 2, \ldots, n\}, \tag{36}$$

where h denotes the binary entropy function. For larger numbers of arguments n, the lower bound in (36) is quite small. In fact, completeness is a rather weak requirement. For instance, the function g: GF(2)$^n \to \{1, -1\}$ defined by

$$g(\mathbf{x}) = \begin{cases} -1 & \text{for all } \mathbf{x}\text{'s with Hamming weight 1,} \\ 1 & \text{for the other } \mathbf{x}\text{'s} \end{cases} \tag{37}$$

is complete, since, for all $i \in \{1, 2, \ldots, n\}$, $g(\mathbf{0}) \neq g(\mathbf{0} \oplus \mathbf{c}_i)$. However, this function is of little cryptographic worth, since it takes in most cases (i.e., $2^n - n$ times) the same value over its domain.

Anyway, (36) underscores an essential difference between completeness and conditional entropies as considered in this paper. Completeness is a characterization of the sensitivity of boolean functions to small input *changes*, whereas conditional entropies measure how strongly the output of a function depends on subsets of its input variables.

In [9] Webster and Tavares define the avalanche effect for an S-box as the property that, on average, half the output bits change when a single input bit is complemented. Clearly, this definition takes two notions into account simultaneously:

- the dependence between input and output of the S-box ("vertical" dependence), and
- the dependence between the various outputs of the S-box ("horizontal" dependence).

In this paper these two notions are considered separately: conditional entropies $H(Y_s | X_{i_1} \cdots X_{i_k})$ deal with the vertical dependence and conditional entropies $H(Y_s | Y_{i_1} \cdots Y_{i_l})$ deal with horizontal dependences. It is therefore difficult to compare formally the avalanche effect to our characterization of S-boxes. Note that the avalanche effect, like completeness, is concerned with sensitivity to small input *changes*.

In the same paper [9] Webster and Tavares also give the following definition.

Definition 2. A boolean function fulfills the strict avalanche criterion (SAC) if its output changes with a probability of one-half whenever a single input bit is complemented.

The SAC of higher order was defined in [2] and expressed in a simpler way in [4].

Definition 3. A boolean function of n variables is said to fulfill the SAC of order m ($1 \leq m \leq n-2$) if and only if any function obtained from f by keeping m of its input bits constant satisfies the SAC (for any choice of the positions and of the values of the m constant bits).

Both the SAC and the higher-order SAC aim at maximizing expressions of the form

$$H(g(\mathbf{x} \oplus \mathbf{c}_i) | g(\mathbf{x})), \tag{38}$$

where g denotes either a boolean function (for the SAC) or a subfunction obtained from a boolean function by keeping a certain number of its input bits constant (higher-order SAC). In other words, the SAC and the higher-order SAC, like the completeness and the avalanche effect, consider the sensitivity of a function to input *changes*. None of these notions can said to be stronger or weaker than the maximization of conditional entropies, since the points of view adopted are different.

The nonlinearity of boolean functions [5] is very important in cryptography. The distance $\delta(f)$ of a function f to the set of affine functions is defined as the minimum of the Hamming distances of f to all the affine functions, and is related to the Walsh transform F of f: $\mathrm{GF}(2)^n \to \{1, -1\}$ according to

$$\delta(f) = 2^{n-1} - \tfrac{1}{2} \max_{w} |F(\mathbf{w})|. \tag{39}$$

Equation (39) shows that functions with a "good degree of nonlinearity" do not have any large Walsh coefficient. In the maximization of conditional entropies, (24) shows that small values of $|F(\mathbf{w})|$ are also required, at least for $\mathbf{w}$'s having Hamming weights smaller than n. But the value of $F(11\cdots1)$ acts upon none of the conditional entropies given 0, 1, 2, ..., n input variables. Consequently, there might be functions with good entropy profiles and very small (or vanishing) distances to affine functions. Consider, for example, $f(x_1, x_2, \ldots, x_n) = x_1 \oplus x_2 \oplus \cdots \oplus x_n$. It is a linear function ($\delta(f) = 0$) which possesses an optimal set of conditional entropies $H(f(\mathbf{X}) | X_{i_1}, X_{i_2}, \ldots, X_{i_k}) = 1$, $\forall k \in \{0, 1, \ldots, n-1\}$. Therefore, a careful design of S-boxes cannot consider exclusively the maximization of conditional entropies. The degree of nonlinearity of the functions involved has to be checked separately, as in step 6 of our algorithm.

3. Statistical Independency of Boolean Functions

In this section we are concerned with the problem of minimizing the mutual information $I(Y_i; \mathbf{Y}')$ between the output random variable Y_i and a random output subvector $\mathbf{Y}'$. This is equivalent to the requirement that any 2, 3, ..., m outputs of the S-box are statistically independent. The statistical independency of boolean functions is investigated in the following.

We consider to boolean functions

$$f_i \,(\text{resp. } f_j)\colon \mathrm{GF}(2)^n \to \{1, -1\},$$
$$\mathbf{X} \mapsto Y_i \,(\text{resp. } Y_j).$$

If their outcomes Y_i and Y_j are looked at as random variables, their statistical independency can be expressed as

$$P_{Y_iY_j}(y_iy_j) = P_{Y_i}(y_i)\cdot P_{Y_j}(y_j), \qquad y_i, y_j \in \{-1, 1\}. \tag{40}$$

If (40) is fulfilled, we say that the *functions* f_i and f_j *are statistically independent*. If we denote by $N_1^{f_i}$ (resp. $N_{-1}^{f_i}$) the number of 1's (resp. of -1's) in the truth table of f_i, we can write

$$\sum_{X \in \mathrm{GF}(2)^n} f_i(\mathbf{X}) = N_1^{f_i} - N_{-1}^{f_i} \tag{41}$$

and

$$2^n = N_1^{f_i} + N_{-1}^{f_i}. \tag{42}$$

Therefore,

$$N_1^{f_i} = 2^{n-1} + \tfrac{1}{2} \sum_{X \in \mathrm{GF}(2)^n} f_i(\mathbf{X}), \tag{43}$$

$$N_{-1}^{f_i} = 2^{n-1} - \tfrac{1}{2} \sum_{X \in \mathrm{GF}(2)^n} f_i(\mathbf{X}), \tag{44}$$

and

$$P_{Y_i}(1) = \tfrac{1}{2}\left(1 + 2^{-n} \sum_{X \in \mathrm{GF}(2)^n} f_i(\mathbf{X})\right), \tag{45}$$

$$P_{Y_i}(-1) = \tfrac{1}{2}\left(1 - 2^{-n} \sum_{X \in \mathrm{GF}(2)^n} f_i(\mathbf{X})\right), \tag{46}$$

i.e.,

$$P_{Y_i}(y_i) = \tfrac{1}{2}\left(1 + y_i\cdot 2^{-n} \sum_{X \in \mathrm{GF}(2)^n} f_i(\mathbf{X})\right). \tag{47}$$

On the other hand,

$$\sum_{X \in \mathrm{GF}(2)^n} f_i(\mathbf{X}) f_j(\mathbf{X}) = \#\{\mathbf{X} : f_i(\mathbf{X}) = f_j(\mathbf{X})\} - \#\{\mathbf{X} : f_i(\mathbf{X}) \neq f_j(\mathbf{X})\} \tag{48}$$

and

$$2^n = \#\{\mathbf{X} : f_i(\mathbf{X}) = f_j(\mathbf{X})\} + \#\{\mathbf{X} : f_i(\mathbf{X}) \neq f_j(\mathbf{X})\}. \tag{49}$$

We denote by $N_{y_i,y_j}^{f_if_j}$ the number of $\mathbf{X} \in \mathrm{GF}(2)^n$ such that $f_i(\mathbf{X}) = y_i$ and $f_j(\mathbf{X}) = y_j$. We get from (48) and (49)

$$N_{1,1}^{f_if_j} + N_{-1,-1}^{f_if_j} = 2^{n-1} + \tfrac{1}{2} \sum_{X \in \mathrm{GF}(2)^n} f_i(\mathbf{X}) f_j(\mathbf{X}), \tag{50}$$

$$N_{1,-1}^{f_if_j} + N_{-1,1}^{f_if_j} = 2^{n-1} - \tfrac{1}{2} \sum_{X \in \mathrm{GF}(2)^n} f_i(\mathbf{X}) f_j(\mathbf{X}). \tag{51}$$

Equations (50) and (51), together with

$$\sum_{X \in \mathrm{GF}(2)^n} f_i(\mathbf{X}) + f_j(\mathbf{X}) = 2N_{1,1}^{f_if_j} - 2N_{-1,-1}^{f_if_j}, \tag{52}$$

$$\sum_{X \in \mathrm{GF}(2)^n} f_i(\mathbf{X}) - f_j(\mathbf{X}) = 2N_{1,-1}^{f_if_j} - 2N_{-1,1}^{f_if_j}, \tag{53}$$

yield

$$N_{1,1}^{f_if_j} = \tfrac{1}{4}\left(2^n + \sum_{X \in \mathrm{GF}(2)^n} f_i(\mathbf{X}) + \sum_{X \in \mathrm{GF}(2)^n} f_j(\mathbf{X}) + \sum_{X \in \mathrm{GF}(2)^n} f_i(\mathbf{X}) f_j(\mathbf{X})\right), \tag{54}$$

$$N^{f_i f_j}_{-1,-1} = \tfrac{1}{4}\Big(2^n - \sum_{X \in GF(2)^n} f_i(\mathbf{X}) - \sum_{X \in GF(2)^n} f_j(\mathbf{X}) + \sum_{X \in GF(2)^n} f_i(\mathbf{X}) f_j(\mathbf{X})\Big), \quad (55)$$

$$N^{f_i f_j}_{1,-1} = \tfrac{1}{4}\Big(2^n + \sum_{X \in GF(2)^n} f_i(\mathbf{X}) - \sum_{X \in GF(2)^n} f_j(\mathbf{X}) - \sum_{X \in GF(2)^n} f_i(\mathbf{X}) f_j(\mathbf{X})\Big), \quad (56)$$

$$N^{f_i f_j}_{-1,1} = \tfrac{1}{4}\Big(2^n - \sum_{X \in GF(2)^n} f_i(\mathbf{X}) + \sum_{X \in GF(2)^n} f_j(\mathbf{X}) - \sum_{X \in GF(2)^n} f_i(\mathbf{X}) f_j(\mathbf{X})\Big), \quad (57)$$

or more generally

$$N^{f_i f_j}_{y_i, y_j} = \tfrac{1}{4}\Big(2^n + y_i \sum_{X \in GF(2)^n} f_i(\mathbf{X}) + y_j \sum_{X \in GF(2)^n} f_j(\mathbf{X}) + y_i y_j \sum_{X \in GF(2)^n} f_i(\mathbf{X}) f_j(\mathbf{X})\Big). \quad (58)$$

Thus

$$P_{Y_i Y_j}(y_i y_j) = 2^{-n} N^{f_i f_j}_{y_i, y_j} \quad (59)$$

$$= 2^{-n-2}\Big(2^n + y_i \sum_{X \in GF(2)^n} f_i(\mathbf{X}) + y_j \sum_{X \in GF(2)^n} f_j(\mathbf{X}) + y_i y_j \sum_{X \in GF(2)^n} f_i(\mathbf{X}) f_j(\mathbf{X})\Big). \quad (60)$$

Equations (47) and (60) inserted in (40) yield after simplification

$$\sum_{X \in GF(2)^n} f_i(\mathbf{X}) f_j(\mathbf{X}) = 2^{-n} \sum_{X \in GF(2)^n} f_i(\mathbf{X}) \cdot \sum_{X \in GF(2)^n} f_j(\mathbf{X}). \quad (61)$$

Equation (61) is a *necessary and sufficient* condition for two functions $f_i(\mathbf{X})$ and $f_j(\mathbf{X})$ to be statistically independent. If we denote by $f_{ij}(\mathbf{X})$ the term-by-term product (bitwise AND) $f_i(\mathbf{X}) \cdot f_j(\mathbf{X})$, by $F_{ij}(\mathbf{w})$ its Walsh transform, and by $F_i(\mathbf{w})$ (resp. $F_j(\mathbf{w})$) the Walsh transform of f_i(resp. f_j), (61) can be written as

$$F_{ij}(\mathbf{0}) = 2^{-n} F_i(\mathbf{0}) F_j(\mathbf{0}), \quad (62)$$

which gives an equivalent way of checking the statistical independency of two boolean functions.

In a similar way we can derive the condition for three boolean functions f_i, f_j, and f_k to be statistically independent, that is,

$$P_{Y_i Y_j Y_k}(y_i y_j y_k) = P_{Y_i}(y_i) \cdot P_{Y_j}(y_j) \cdot P_{Y_k}(y_k), \qquad y_i, y_j, y_k \in \{-1, 1\}. \quad (63)$$

Assuming that f_i, f_j, and f_k are pairwise statistically independent, (63) holds under the following necessary and sufficient condition:

$$\sum_{X \in GF(2)^n} f_i(\mathbf{X}) f_j(\mathbf{X}) f_k(\mathbf{X}) = 2^{-2n} \sum_{X \ldots} f_i(\mathbf{X}) \cdot \sum_{X \ldots} f_j(\mathbf{X}) \cdot \sum_{X \ldots} f_k(\mathbf{X}). \quad (64)$$

For four functions f_i, f_j, f_k, and f_l that are pairwise and tripletwise statistically independent, the additional condition for overall statistical independency is

$$\sum_{X \in GF(2)^n} f_i(\mathbf{X}) f_j(\mathbf{X}) f_k(\mathbf{X}) f_l(\mathbf{X}) = 2^{-3n} \sum_{X \ldots} f_i(\mathbf{X}) \cdot \sum_{X \ldots} f_j(\mathbf{X}) \cdot \sum_{X \ldots} f_k(\mathbf{X}) \cdot \sum_{X \ldots} f_l(\mathbf{X}). \quad (65)$$

Due to the similarity of equations (61), (63), (64), (65) and their derivations, we state the following generalization:

$m-1/1$-valued boolean functions $f_{i_1}, f_{i_2}, \ldots, f_{i_m}$ are pairwise, tripletwise, ..., m-tuplewise statistically independent if and only if, for $s = 2, 3, \ldots, m$ and for any combination of distinct $j_1, j_2, \ldots, j_s \in \{i_i, i_2, \ldots, i_m\}$,

$$\sum_{X \in GF(2)^n} f_{j_1}(\mathbf{X}) f_{j_2}(\mathbf{X}) \cdots f_{j_s}(\mathbf{X}) = 2^{-(s-1)n} \sum_{X \ldots} f_{j_1}(\mathbf{X}) \cdot \sum_{X \ldots} f_{j_2}(\mathbf{X}) \cdots \sum_{X \ldots} f_{j_s}(\mathbf{X}) \quad (66)$$

or

$$F_{j_1 j_2 \cdots j_s}(\mathbf{0}) = 2^{-(s-1)n} F_{j_1}(\mathbf{0}) F_{j_2}(\mathbf{0}) \cdots F_{j_s}(\mathbf{0}), \quad (67)$$

using the notation introduced in (62). It can easily be checked that each of the 8 DES S-boxes has the property that its four outputs are pairwise, tripletwise, and quadrupletwise statistically independent, or, in other words,

$$H(Y_i | Y_j) = H(Y_i | Y_j Y_k) = H(Y_i | Y_j Y_k Y_l) = H(Y_i) \text{ for distinct } i, j, k, l \in \{1, 2, 3, 4\}. \quad (68)$$

4. Conclusions

Two categories of results are presented in this paper: some general, mathematical results and some results aiming at the design of cryptographically strong S-boxes. The general mathematical results consist of:

- An expression of the conditional entropies of the result of a boolean function (considered as a random variable), conditioned on k of its random, uniformly distributed arguments. These conditional entropies are shown to depend only on the values of the Walsh transform of the function calculated for arguments $\mathbf{w}$ with $H(\mathbf{w}) \leq k$.
- Necessary and sufficient conditions for the statistical independency of the results of two, three, ... boolean functions (again considered as random variables).

The results aiming at the design of S-boxes comprise:

- The definition of the entropy profile of a boolean function. For a boolean function to be "cryptographically strong," we assume that its entropy profile should exhibit no "deep canyon" and that it should lie entirely above some chosen curve.
- An algorithm to obtain boolean functions with "good" entropy profiles by successive refinement of randomly chosen functions.

References

[1] H. Feistel. Cryptography and computer privacy. *Scientific American*, **228**(5): 15–23, May 1973.

[2] R. Forré. The strict avalanche criterion: spectral properties of boolean functions and an extended definition. In *Advances in Cryptology: Crypto '88 Proceedings*, Springer-Verlag, Berlin, 1990, pp. 450–468.

[3] J. B. Kam and G. I. Davida. Structured design of substitution-permutation encryption networks. *IEEE Transactions on Computers*, **28**(10): 747–753, October 1979.

[4] S. Lloyd. Counting functions satisfying a higher-order strict avalanche criterion. In *Advances on Cryptology: Eurocrypt '89 Proceedings*, Springer-Verlag, Berlin, to appear.

[5] W. Meier and O. Staffelbach. Nonlinearity criteria for cryptographic functions. In *Advances in Cryptology: Eurocrypt '89 Proceedings*, Springer-Verlag, Berlin, to appear.

[6] O. S. Rothaus. On bent functions. *Journal of Combinatorial Theory (A)*, **20**: 300–305, 1976.

[7] T. Siegenthaler. Correlation-immunity of nonlinear combining functions for cryptographic applications. *IEEE Transactions on Information Theory*, **30**: 776–780, October 1984.

[8] R. C. Titsworth. Correlation Properties of Cyclic Sequences. Ph.D. thesis, California Institute of Technology, Pasadena, California, 1962.

[9] A. F. Webster and S. E. Tavares. On the design of S-boxes. In *Advances in Cryptology: Crypto '85 Proceedings*, Springer-Verlag, Berlin, 1986.

[10] G. Z. Xiao and J. L. Massey. A spectral chracterization of correlation-immune combining functions. *IEEE Transactions on Information Theory*, **34**(3): 569–571, May 1988.