Journal of Universal Computer Science, vol. 4, no. 8 (1998), 705-717
submitted: 9/12/97, accepted: 4/8/98, appeared: 28/8/98

On Cryptographic Properties of Random Boolean Functions

Daniel Olejár
Department of Computer Science
Comenius University
olejar@dcs.fmph.uniba.sk

Martin Stanek
Department of Computer Science
Comenius University
stanek@dcs.fmph.uniba.sk

Abstract: Boolean functions used in cryptographic applications have to satisfy various cryptographic criteria. Although the choice of the criteria depends on the cryptosystem in which they are used, there are some properties (balancedness, nonlinearity, high algebraic degree, correlation immunity, propagation criteria) which a cryptographically strong Boolean function ought to have. We study the above mentioned properties in the set of all Boolean functions (all balanced Boolean functions) and prove that almost every Boolean function (almost every balanced Boolean function) satisfies all above mentioned criteria on levels very close to optimal and therefore can be considered to be cryptographically strong.

1 Introduction

The robustness of a cryptosystem substantially depends on its underlying elements. Since Boolean functions are frequently used in various cryptosystems, it would be interesting to determine the properties which cryptographically strong Boolean functions should to have and to find methods how to construct them. Nevertheless, none of these problems can be solved in general. Boolean functions meeting all possible cryptographic requirements on the highest possible level do not exist (e.g. balancedness exludes maximal nonlinearity, etc.). Much of cryptographic research in Boolean functions has therefore concentrated on Boolean functions satisfying criteria formulated by the designers and cryptanalysts of real-world cryptosystems [see Adams, Tavares 90, Camion, Carlet, Charpin, Sendrier 91, Millan, Clark, Dawson 97, Seberry 93a, 93b]. They studied various cryptographically desirable properties of Boolean functions, the mutual relations of this properties and construction methods. The following cryptographic properties of Boolean functions belong to the most important and most frequently studied: balancedness, nonlinearity, satisfying propagation criteria, especially SAC (Strict Avalanche Criterion), the absence of linear structures, high algebraic degree, correlation immunity, etc. Boolean functions satisfying some of these criteria are considered to be cryptographically strong. Formally: let $P = (p_1, \ldots, p_m)$ be a set of properties of Boolean functions expressed as real-valued parameters; let $\Lambda = (\lambda_1, \ldots, \lambda_m)$ denote the set of required levels of the properties P. The Boolean function f is said to be *cryptographically strong on the level Λ with respect to the properties P*, if $p_i(f) \geq \lambda_i,\ i = 1, \ldots, m$; otherwise, f is *cryptographically weak on the level Λ with respect to P*. The choice

of properties P and setting the level Λ depends on the cryptosystem itself, on its intended use and on the state-of-art of cryptanalysis. By the proper choice of P and Λ the designer can shield his cryptosystem against known cryptanalytic attacks based on the underlying Boolean functions, but he cannot guarantee the robustness of the cryptosystem against unknown cryptanalytic attacks (or against attacks using some other weak points of the cryptosystem). We study Boolean functions with respect to the above mentioned criteria. The cryptographically strong Boolean function in this paper means that it satisfies the criteria at levels asymptotically equal to optimal.

There is no general method known (except for the full search) of how to construct Boolean functions satisfying an arbitrarily chosen subset of cryptographic criteria. Since the full search is limited to Boolean functions with at most five variables and real world applications require Boolean functions with more variables, it is of interest be interesting to know how cryptographic properties are distributed in the set of all n-ary Boolean functions.

Mitchell [see Mitchell 90] studied Boolean functions satisfying various cryptographic criteria such as balancedness, nonlinearity, nondegenerancy, correlation immunity and symmetry. He enumerated or estimated the cardinality of classes of Boolean functions satisfying various combinations of the above mentioned criteria. On the other hand he considered the cryptographic criteria as qualitative properties and did not distinguish, e.g. between strong and weak nonlinearity. Moreover, he concentrated his attention to criteria that are combinatoricaly tractable, although their cryptographic value is at least questionable (symmetry) or to weaker criteria that follow from stronger ones (nondegenerancy — SAC).

We adopt a different approach — we chose the most important cryptographic properties of Boolean functions, consider them as quantitative parameters and estimate how many Boolean functions satisfy the particular criterion on some level or of some order. We prove that almost every Boolean function satisfies the set of the most important cryptographic criteria at the suboptimal — but asymptotically optimal — level.

2 Preliminaries

An *n-ary Boolean function* is a mapping $f : \{0,1\}^n \to \{0,1\}$. The set (class) of all n-ary Boolean functions will be denoted by the symbol P_2^n. We shall consider only n-ary Boolean functions in this paper and therefore the notion Boolean function stands for n-ary Boolean function (if not otherwise stated). Let M_n be a class of n-ary Boolean functions, the symbol $|M_n|$ denotes the cardinality of M_n. Let $\mathcal{P}$ be a property of Boolean functions and let $\mathcal{P}(M_n)$ denote the subclass of Boolean functions, $\mathcal{P}(M_n) \subseteq M_n$, satisfying $\mathcal{P}$. We say that *Boolean function from M_n has property $\mathcal{P}$ almost surely*, if

$$\lim_{n\to\infty} |\mathcal{P}(M_n)|/|M_n| = 1.$$

An n-ary Boolean function f will be described by its *truth table:* $f(x_1,\ldots,x_n) = (f(0,\ldots,0), f(0,\ldots,0,1),\ldots,f(1,\ldots,1))$. The truth table of an n-ary Boolean function f is a binary vector of length 2^n and will be denoted by $\mathrm{tt}(f)$. Let $\alpha = (a_1,\ldots,a_m), \beta = (b_1,\ldots,b_m)$ be two binary vectors of length m. The symbol

$\alpha \oplus \beta$ denotes the bitwise XOR-operation of α, β and the symbol $\langle \alpha, \beta \rangle$ denotes the inner product of the vectors α and β, i.e. $\langle \alpha, \beta \rangle = a_1 b_1 \oplus a_2 b_2 \oplus \ldots \oplus a_m b_m$. The *Hamming weight* of a binary vector α, denoted $wt(\alpha)$, is the number of ones in α. The Hamming weight of a Boolean function f, $\mathrm{wt}(\mathrm{tt}(f))$, will be denoted by $\mathrm{wt}(f)$. Let f be an n-ary Boolean function. Then f is a *balanced Boolean function* if $\mathrm{wt}(f) = 2^{n-1}$. The set of all balanced Boolean functions from P_2^n will be denoted by Bal_n. The *Hamming distance of vectors* α, β, denoted $d(\alpha, \beta) = \mathrm{wt}(\alpha \oplus \beta)$, is the number of bits in which α and β differ. Let f, g be two n-ary Boolean functions. The symbol $d(f, g)$ denotes the *Hamming distance of functions* f, g; $d(f, g) = d(\mathrm{tt}(f), \mathrm{tt}(g))$. Let $f(x_1, \ldots, x_n)$ be an n-ary Boolean function. *The algebraic normal form* (ANF) of f is given by the following representation of f for each $x_1, \ldots, x_n \in \{0, 1\}$:

$$f(x_1, \ldots, x_n) = a_0 \oplus a_1 x_1 \oplus a_2 x_2 \oplus a_3 x_2 x_1 \oplus \ldots \oplus a_{2^n-1} x_n \ldots x_1, \qquad (1)$$

where $a_i \in \{0, 1\}, i = 0, 1, \ldots, 2^n - 1$, is a constant. The Boolean function $f(x_1, \ldots, x_n)$ is said to be *affine* if its ANF contains only linear terms:

$$f(x_1, \ldots, x_n) = a_0 \oplus a_1 x_1 \oplus a_2 x_2 \oplus a_4 x_3 \oplus \ldots \oplus a_{2^{n-1}} x_n,$$

where $a_j \in \{0, 1\}, j = 0, 1, 2, 4, \ldots, 2^{n-1}$. An ANF will be represented by the vector of its coefficients $(a_0, a_1, \ldots, a_{2^n-1})$ — the *ANF-vector*. An affine Boolean function is said to be *linear*, if the absolute term a_0 in its ANF is equal to 0. Let $f(x_1, \ldots, x_n)$ be an n-ary Boolean function. The *algebraic degree* of f is

$$\deg(f) = \max_j \{\mathrm{wt}(j) |\ a_j = 1\},$$

where the parameter j is represented in binary and ranges over all indices of ANF f; $j = 0, \ldots, 2^n - 1$. The algebraic degree of the Boolean function f corresponds to the maximal length (number of variables) of a conjunction in its ANF.

We will use the following geometrical model of P_2^n. The *Boolean hypercube*, B^{2^n}, is a labelled graph containing 2^{2^n} vertices $v_0, \ldots, v_{2^{2^n}-1}$. Each vertex of B^{2^n} is labelled by a binary vector of length 2^n. Two vertices u, v of B^{2^n} are adjacent, if the corresponding vectors α, β differ in one bit: $d(\alpha, \beta) = 1$. As can be easily seen, B^{2^n} represents P_2^n; every function from P_2^n corresponds to a (unique) vertex in B^{2^n} and two Boolean functions differing in one component in their truth tables, correspond to adjacent vertices in B^{2^n}. Therefore, the vertices of B^{2^n} will be considered in the following as n-ary Boolean functions.

Let B^{2^n} be a hypercube and let α be a vertex of B^{2^n}. The subgraph $S(2^n, \alpha, r)$ of B^{2^n} induced by the set of vertices

$$\{\beta \mid d(\alpha, \beta) \leq r\}$$

is traditionally called *the sphere of B^{2^n} with centre α and diameter r.* Since we are not interested in connectivity and any similar problems, we do not distinguish between the graph $S(2^n, \alpha, r)$ and its vertex set.

Boolean functions with various (cryptographic) properties form subgraphs of the Boolean hypercube. To estimate the size of these subgraphs, we need a precise estimate on the number of vertices of the sphere with diameter r.

The following asymptotic estimate of binomial coefficient was proved by Knuth et al. [see Graham, Knuth, Patashnik 94].

Lemma 1. *Let $0 < \varepsilon < 1/6$ be a constant, let $|k| \leq 2^{(n-1)(1/2+\varepsilon)}$ be a positive integer. Then*

$$\binom{2^n}{2^{n-1}-k} = \frac{2^{2^n+1/2}}{\sqrt{\pi \cdot 2^n}} \cdot e^{-k^2/2^{n-1}} \left(1 + O(2^{n(3\varepsilon-1/2)})\right). \tag{2}$$

The cardinality of $S(2^n, \alpha, r)$ will be expressed by means of the binomial coefficient $\binom{2^n}{2^{n-1}}$. Its value can be obtained (as a special case for $k = 0$) from (2) but a more precise bound directly follows from Stirling's formula.

Lemma 2. *Let n be a positive integer. Then it holds*

$$\binom{2^n}{2^{n-1}} = \frac{2^{2^n}}{\sqrt{\pi \cdot 2^{n-1}}} \left(1 - O(2^{-n})\right). \tag{3}$$

3 Estimating the size of spheres in the Boolean hypercube

In this section we estimate the cardinality of the sphere $S(2^n, \alpha, r)$ and find two important values of r. Further results of this paper are immediate applications of the bounds constructed in this section.

Let for simplicity $s(n, r) = |S(2^n, \alpha, r)|$, where α is an arbitrary vertex of B^{2^n}.

Theorem 3. *Let $r_1 = 2^{n-1} - c_1\sqrt{n} \cdot 2^{n/2}$; $r_2 = 2^{n-1} - c_1\sqrt{\lg n} \cdot 2^{n/2}$, where $c_1 = \sqrt{(1+\varepsilon_0)(1/2)\ln 2}$ and $\varepsilon_0 > 0$ is an arbitrary constant. Then we have*

$$s(n, r_1) = 2^{2^n - n} \cdot O(2^{-n\varepsilon_0}) \tag{4}$$

and

$$s(n, r_2) = O(2^{2^n}/n^{1+\varepsilon_0}) \tag{5}$$

Proof. We divide the sum $\sum_{0 \leq k \leq r_1} \binom{2^n}{k}$ into two sums Σ_1, Σ_2 and estimate them separately. Let $m_0 = 2\sqrt{n} \cdot 2^{n/2}$, then

$$\sum_{0 \leq k \leq r_1} \binom{2^n}{k} = \sum_{0 \leq k \leq 2^{n-1} - m_0} \binom{2^n}{k} + \sum_{2^{n-1} - m_0 < k \leq r_1} \binom{2^n}{k}. \tag{6}$$

Since the sequence $\binom{2^n}{k}$ is unimodal, the first sum, Σ_1, can be bounded by the product of its last (largest) term and the upper bound of the number of its terms.

$$\Sigma_1 < \binom{2^n}{2^{n-1} - m_0} \cdot 2^n = 2^{2^n} \cdot O\left(\left(\frac{\sqrt{2}}{e^8}\right)^n\right) = o(2^{2^n - n}), \tag{7}$$

since $\frac{\sqrt{2}}{e^8} < \frac{1}{2}$. To estimate the second sum (Σ_2) of (6), we change the order of summation and then use (2) to estimate the summand:

$$\Sigma_2 = \sum_{c_1\sqrt{n}\cdot 2^{n/2} \le j < 2\sqrt{n}\cdot 2^{n/2}} \binom{2^n}{2^{n-1}-j}$$
$$= \frac{2^{2^n+1/2}}{\sqrt{\pi\cdot 2^n}}\left(1+O(2^{n(3\varepsilon-1/2)})\right)\cdot \sum_{c_1\sqrt{n}\cdot 2^{n/2} \le j < 2\sqrt{n}\cdot 2^{n/2}} e^{-j^2/2^{n-1}}.$$

Now we transform the summation range by setting $k = j - c_1\sqrt{n}\cdot 2^{n/2}$ and then concentrate our effort on constructing an upper bound on the sum which appears in (8) and will be denoted by Σ_3:

$$\Sigma_2 = \frac{2^{2^n+1/2}}{\sqrt{\pi\cdot 2^n}}\left(1+O(2^{n(3\varepsilon-1/2)})\right)\cdot \sum_{0\le k<(2-c_1)\sqrt{n}\cdot 2^{n/2}} \exp\left[\frac{-(c_1\sqrt{n}\cdot 2^{n/2}+k)^2}{2^{n-1}}\right] \tag{8}$$

$$\Sigma_3 = e^{-2c_1^2 n}\cdot \sum_{0\le k<(2-c_1)\sqrt{n}\cdot 2^{n/2}} \exp\left[\frac{-2c_1\sqrt{n}\cdot k}{2^{n/2-1}}\right]\cdot \exp\left[\frac{-k^2}{2^{n-1}}\right]$$
$$< e^{-2c_1^2 n}\cdot \sum_{0\le k<(2-c_1)\sqrt{n}2^{n/2}} \exp\left[\frac{-k^2}{2^{n-1}}\right]. \tag{9}$$

The sum (denoted by Σ_4) in (9) can be bounded by the integral

$$\Sigma_4 < 1 + \int_0^{(2-c_1)\sqrt{n}2^{n/2}} e^{-x^2/2^{n-1}}dx,$$

which can be expressed by the distribution function of normal distribution. Let $u^2/2 = x^2/2^{n-1}$, then

$$\Sigma_4 < 1 + 2^{n/2-1}\sqrt{2\pi}\int_0^{(2-c_1)\sqrt{n}\cdot 2} e^{-u^2/2}du = O(2^{n/2}). \tag{10}$$

And

$$\Sigma_2 = 2^{2^n-n}\cdot O(2^{-n\varepsilon_0}), \tag{11}$$

where ε_0 is a positive constant. Taking into account (11) and (7) we obtain (4).

To prove (5), it is sufficient to estimate the value $\Sigma_5 = s(n,r_2) - s(n,r_1)$:

$$\Sigma_5 = \sum_{c_1\sqrt{\lg n}2^{n/2} \le k < c_1\sqrt{n}2^{n/2}} \binom{2^n}{2^{n-1}-k}$$
$$= 2^{2^n-n/2}\cdot O(\sum_{c_1\sqrt{\lg n}2^{n/2} \le k < c_1\sqrt{n}2^{n/2}} e^{-k^2/2^{n-1}}). \tag{12}$$

We can proceed in the same way as in the previous case. Let Σ_6 denote the last sum in (12). Following the steps (9) and (10) we obtain

$$\Sigma_6 = e^{-2c_1^2 \lg n} \cdot O(2^{n/2}) = e^{-(1+\varepsilon_0)\ln 2 \cdot \lg n} \cdot O(2^{n/2}) = n^{-(1+\varepsilon_0)} \cdot O(2^{n/2}).$$

Therefore

$$\Sigma_5 = 2^{2^n} \cdot O(n^{-(1+\varepsilon_0)}),$$

and

$$s(n, r_2) = 2^{2^n - n} \cdot O(2^{-n\varepsilon_0}) + 2^{2^n} \cdot O(n^{-(1+\varepsilon_0)}) = 2^{2^n} \cdot O(n^{-(1+\varepsilon_0)}).$$

□

We shall pay special attention to balanced Boolean functions because of their significance in cryptographic design. To describe the cryptographic properties of an average n-ary balanced Boolean function, we need to estimate the number of balanced vectors (vertices) in $S(2^n, \alpha, r)$, with the centre $\alpha \in \mathrm{Bal}_n$ (where Bal_n denotes the set of all balanced Boolean functions of n variables). To simplify the notation, let $s_b(n, r)$ denote the value in question. Let α be the vector (truth table) of a balanced n-ary Boolean function. Without loss of generality we can assume that $\alpha = (1, 1, \ldots 1, 0, 0, \ldots, 0)$. There are no balanced neighbouring vectors, nor balanced vectors lying at odd distance from the vector-vertex α in B^{2^n}. Therefore we count the number of balanced vectors/vertices lying at distance $2k$. Such a vector can be obtained from α by replacing k bits from the first half of α by zeroes and (to save the balancedness of the constructed vector) replacing k zeroes from the second half of α by ones. Therefore the number of vertices corresponding to balanced Boolean functions lying at distance $2r$ or smaller from α is

$$s_b(n, 2r) = \sum_{k=0}^{r} \binom{2^{n-1}}{k}^2 .$$

Analogously as in the previous case, we estimate the value $s_b(n, r)$ and find two important values of the parameter r.

Theorem 4. *Let* $r_3 = 2^{n-1} - c_3 \cdot 2^{n/2+1}\sqrt{n}$ *and* $r_4 = 2^{n-1} - c_4 \cdot 2^{n/2+1}\sqrt{\lg n}$, *where* $c_3 = \sqrt{(1+\varepsilon_0)(\ln 2)/8}, c_4 = \sqrt{(1+\varepsilon_0)/(8 \lg e)}$ *and* ε_0 *is an arbitrary positive constant. Then*

$$s_b(n, r_3) = 2^{2^n - 3n/2} \cdot O(2^{-\varepsilon_0 n}) \tag{13}$$

$$s_b(n, r_4) = 2^{2^n - n/2} \cdot O(1/n^{1+\varepsilon_0}) \tag{14}$$

Proof. The proof of Theorem 4 is similar to the proof of Theorem 3 and therefore omitted. □

Remark. In the rest of this paper the symbols c_1, c_3, c_4; r_1, r_2, r_3, r_4 denote constants and values of parameter r derived in Theorems 3 and 4.

4 Balancedness of Boolean functions

Balancedness is one of the most important cryptographic properties of Boolean functions. Bijective S-boxes are created from balanced Boolean functions and the equiprobability of characters of the output alphabet is the basic condition of a cryptographically strong cryptosystem.

As can be easily seen, $|\,\mathrm{Bal}_n\,| = \binom{2^n}{2^{n-1}}$. Lemma 2 states that the number of balanced Boolean functions in P_2^n is negligible. On the other hand, almost every Boolean function is "almost balanced".

Theorem 5. *Let f be an n-ary Boolean function, $\phi(n)$ be an arbitrary function such that $\phi(n) \to \infty$ as $n \to \infty$ and let $p \in (0,1)$. Then*

$$p \cdot 2^n - 2^{n/2}\phi(n) < \mathrm{wt}(f) < p \cdot 2^n + 2^{n/2}\phi(n), \tag{15}$$

almost surely.

Proof. The function wt can be considered as a random variable on P_2^n. Let $0 \le k \le 2^n$, then

$$\Pr(\mathrm{wt}(f) = k) = p^k(1-p)^{2^n-k}.$$

The random variable wt has binomial distribution with parameters $2^n, p$. Let $p = 1/2$. The inequalities (15) follow from Chebyshev's inequality. □

5 Nonlinearity of random Boolean functions

Affine and linear Boolean functions play a peculiar role in cryptography. They are cryptographically weak to be directly used for construction of cryptosystems, since linear cryptosystem can be easily broken by solving the system of linear equations. On the other hand, affine Boolean functions are used in various constructions of cryptographically very strong Boolean functions. The (non)linearity of Boolean functions is a qualitative property; to express the measure of nonlinearity, we shall use the following definition [see Pieprzyk, Finkelstein 88]. For any Boolean function f, define

$$N_f = \min_l\{d(f,l)\},$$

where l is an arbitrary affine Boolean function. Obviously, the nonlinearity of an affine Boolean function is zero; the maximal value of the parameter N_f is [see Seberry, Zhang, Zheng 93a]

$$N_f \le 2^{n-1} - 2^{n/2-1}. \tag{16}$$

The nonlinearity of balanced Boolean functions is below the maximal value (16). The following bounds can be found in [Seberry, Zhang, Zeng 93a]:

$$N_f \le \begin{cases} 2^{n-1} - 2^{\frac{1}{2}n-1} - 2, & n \text{ even} \\ \lfloor\lfloor 2^{n-1} - 2^{\frac{1}{2}n-1}\rfloor\rfloor, & n \text{ odd} \end{cases}$$

where $\lfloor\lfloor x\rfloor\rfloor$ denotes the maximum even integer less than or equal to x. We prove lower bounds on the nonlinearity of almost all n-ary (balanced) Boolean functions.

Theorem 6. *1. Let f be an n-ary Boolean function and let g be an n-ary balanced Boolean function. Then*

$$N_f \geq 2^{n-1} - c_1\sqrt{n}2^{n/2}, \tag{17}$$

$$N_g \geq 2^{n-1} - c_3\sqrt{n}2^{n/2+1}, \tag{18}$$

almost surely.

Proof. Let $r \leq r_1$ [see remark below Theorem 4]. If the spheres with centres in affine functions were disjoint, they would contain $\leq 2^{n+1} \cdot s(n, r_1) = 2^{2^n} \cdot O(2^{-n\varepsilon_0})$ vertices. Analogously, 2^{n+1} disjoint spheres with diameter $r \leq r_3$ contain at least $2^{n+1} \cdot s_b(n, r_3) = 2^{2^n - n/2} \cdot O(2^{-n\varepsilon_0})$ balanced Boolean functions. This is negligible with respect to the number of all n-ary balanced Boolean functions. □

6 Correlation immunity

Boolean functions are sometimes used as nonlinear filters of cryptosystems consisting of some linear feedback shift registers (LFSRs). If the filtering function leaks some information on one of its input bits (the output of a LFSR), a successful cryptanalytic attack can be mounted based on this weakness. To avoid the *correlation attack* the filtering function has to be *correlation immune*, [see Siegenthaler 84, Seberry, Zhang, Zheng 93b].

Definition 7. An n-ary Boolean function f is correlation immune of order k, denoted CI_k, if and only if the function $f(x) \oplus \langle x, \alpha \rangle$ is balanced for every $\alpha \in \{0,1\}^n, 1 \leq \mathrm{wt}(\alpha) \leq k$.

In other words, f is CI_k iff the Hamming distance between f and any non-constant affine function l depending on less than or equal to k variables, is exactly 2^{n-1}. Therefore correlation immunity is a qualitative property. We introduce another measure of correlation immunity—*counted correlation characteristic*—*CCC*.

Definition 8. Let f be an n-ary Boolean function. The counted correlation characteristic of order k of the Boolean function f is

$$CCC_k(f) = \min_{\substack{\alpha; 1 \leq \mathrm{wt}(\alpha) \leq k \\ a \in \{0,1\}}} \{\mathrm{wt}(g_{\alpha,a})\},$$

where for fixed $\alpha \in \{0,1\}^n$ and $a \in \{0,1\}$, $g_{\alpha,a}$ is the Boolean function defined by $g_{\alpha,a}(x) = f(x) \oplus \langle x, \alpha \rangle \oplus a$ for each $x \in \{0,1\}^n$.

As can be easily seen, if $CCC_k(f) = 2^{n-1}$ then the function f satisfies CI_k (and vice versa). We estimate the typical value of CCC_k. This is almost the same problem as was studied before (the nonlinearity of an average Boolean function) since we have to estimate the number of Boolean functions lying in spheres with centres in some (chosen) affine functions. Therefore the nonlinearity of a random Boolean function f provides a lower bound on the counted correlation characteristic of f, too. Thus we have from Theorem 6.

Theorem 9. *The counted correlation characteristic of order k of an n-ary Boolean function f satisfies the following inequality almost surely*

$$CCC_k(f) \geq 2^{n-1} - d_k \cdot 2^{n/2}\sqrt{n}, \tag{19}$$

where $1 \leq k \leq n$, and d_k is a positive constant depending on k.

Remark. Though the strict correlation immunity is a rather restrictive property, almost all Boolean functions are "almost correlation immune". The order k of CCC_k does not influence the value of CCC_k substantially. Let us estimate CCC_1 of a random Boolean function. Taking into account the fact that there are $2n$ affine Boolean functions depending on 1 variable and using the upper bounds (5) and (13) we have almost surely

$$CCC_1 \geq 2^{n-1} - c_1\sqrt{\lg n} \cdot 2^{n/2};$$

and for balanced functions

$$CCC_1 \geq 2^{n-1} - c_4\sqrt{\lg n} \cdot 2^{n/2+1};$$

where c_1, c_4 are constants defined in Theorems 3 and 4.

7 Propagation characteristics

Boolean functions used in cryptographic applications have to be very sensitive to small changes of their inputs. That means, if the input value of a Boolean function f is changed, its output value would change with probability 1/2, too. More precisely, we have the following definition.

Definition 10. An n-ary Boolean function f satisfies the *propagation criterion of order k* (PC_k), if

$$\mathrm{wt}(f(x) \oplus f(x \oplus \alpha)) = 2^{n-1}, \tag{20}$$

for each $\alpha \in \{0,1\}^n,\ 1 \leq \mathrm{wt}(\alpha) \leq k$.
The case $k = 1$ is of special importance and is referred to as the *Strict Avalanche Criterion (SAC)* and was introduced by Webster and Tavares [see Webster, Tavares 86].

The enumeration of the set of (n-ary) Boolean functions satisfying PC_k is a very hard combinatorial problem. Tavares [see Tavares 96] presented the following asymptotic bound (constructed by Daniel Biss in 1996) on Pragocrypt '96

$$SAC(n) \sim \frac{2^{2^n - n^2/2 + n}}{\pi^{n/2}}.$$

That means, the average Boolean function does not satisfy SAC. On the other hand, if we replace the strict condition of balancedness in (20) by near-balancedness, we obtain a large set of Boolean functions which are still strong enough for cryptographic applications.

Definition 11. Let f be an n-ary Boolean function. The *counted propagation characteristic of order* k of the Boolean function f is

$$CPC_k(f) = \min_{\alpha;\, 1 \le \mathrm{wt}(\alpha) \le k} \{\mathrm{wt}(g_\alpha)\}\,, \tag{21}$$

where for fixed $\alpha \in \{0,1\}^n$, g_α is the Boolean function defined by $g_\alpha(x) = f(x) \oplus f(x \oplus \alpha)$ for each $x \in \{0,1\}^n$.

As can be easily seen, if f meets PC_k, then $CPC_k(f) = 2^{n-1}$. Now we concentrate on the CPC_1 and construct an upper bound on the number of all n-ary Boolean functions with $CPC_1 \le r$. Since $CPC_k, k = 1, \ldots, n$, is always even, we consider only even values of r. Let $G_{n,r,\alpha}$ (respectively, $G_{n,\le r,\alpha}$) denote the set of all n-ary Boolean functions g satisfying $\mathrm{wt}(g_\alpha) = r$ (respectively, $\mathrm{wt}(g_\alpha) \le r$), where for fixed $\alpha \in \{0,1\}^n$, g_α is the Boolean function defined by $g_\alpha(x) = g(x) \oplus g(x \oplus \alpha)$ for each $x \in \{0,1\}^n$. Let

$$G_{n,\le r,m} = \bigcup_{\alpha;\ 1 \le \mathrm{wt}(\alpha) \le m} G_{n,\le r,\alpha} \qquad \text{and} \qquad G_{n,r,m} = \bigcup_{\alpha;\ \mathrm{wt}(\alpha) = m} G_{n,r,\alpha}\,.$$

We estimate $|G_{n,2k,1}|$. Let $g \in G_{n,2k,1}$. There exists a vector $\beta \in \{0,1\}^n$ $(\mathrm{wt}(\beta) = 1)$, such that $\mathrm{wt}(g(x) \oplus g(x \oplus \beta)) = 2k$ (and for an arbitrary vector $\gamma \in \{0,1\}^n$ $(\mathrm{wt}(\gamma) = 1)$: $\mathrm{wt}(g(x) \oplus g(x \oplus \gamma)) \ge 2k$). Without loss of generality we assume that $\beta = (1,0,\ldots,0)$. Let $(g_0, g_1, \ldots, g_{2^n-1})$ be the truth table of g. Since $g \in G_{n,2k,\beta}$, there exists a k-set I_k of indices; $I_k = \{i_1, \ldots, i_k\}$ where $i_j \in \{0, \ldots, 2^{n-1} - 1\}$ for $j = 1, \ldots, k$ such that

$$g_i = \begin{cases} g_{i+2^{n-1}} \oplus 1 & \text{if } i \in I_k\,; \\ g_{i+2^{n-1}} & \text{else.} \end{cases}$$

The index set I_k can be chosen in $\binom{2^{n-1}}{k}$ ways and there are $2^{2^{n-1}}$ ways how to choose the values of $g_i,\ i = 0, \ldots, 2^{n-1} - 1$. Therefore

$$|G_{n,2k,\beta}| = 2^{2^{n-1}} \cdot \binom{2^{n-1}}{k}\,.$$

If $CPC_1(g) = 2k$, then obviously

$$g \in \bigcup_{\gamma;\ \mathrm{wt}(\gamma)=1} G_{n,2k,\gamma},$$

and therefore the number of g's satisfying $CPC_1(g) = 2k$ does not exceed n times $|G_{n,2k,\beta}|$. Now we can estimate the value of CPC_1 of a random Boolean function.

Theorem 12. *Let f be an n-ary Boolean function. Then*

$$CPC_1(f) \ge 2^{n-1} - 2^{(n+1)/2} \cdot c_1 \cdot \sqrt{\lg(n-1)}$$

almost surely.

Proof. Let $r = 2^{n-1} - 2^{(n+1)/2} c_1 \cdot \sqrt{\lg(n-1)}$. We prove that $|G_{n,\le r,1}| = o(2^{2^n})$. Since

$$G_{n,\le r,1} = \bigcup_{\gamma;\ \mathrm{wt}(\gamma)=1} G_{n,\le r,\gamma},$$

we have from the remark preceding Theorem 12 that:

$$|G_{n,\le r,1}| = |\bigcup_{\gamma;\ \mathrm{wt}(\gamma)=1} G_{n,\le r,\gamma}| \le n \cdot 2^{2^{n-1}} \cdot \sum_{0\le k\le r/2} \binom{2^{n-1}}{k}.$$

To estimate the sum, we use (5) (Theorem 3):

$$|G_{n,\le r,1}| = n \cdot 2^{2^{n-1}} \cdot O\left(\frac{2^{2^{n-1}}}{(n-1)^{1+\varepsilon_0}}\right) = O(2^{2^n}/n^{\varepsilon_0}) = o(2^{2^n}).$$

The theorem follows. □

Remark. If we need to find the lower bound of CPC_q (for $q \in \{1,\ldots,n\}$) of an "average" Boolean function, we have to find a maximal r such that

$$\left[\sum_{0\le i\le q} \binom{n}{i}\right] \cdot 2^{2^n} \cdot \sum_{0\le k\le r/2} \binom{2^{n-1}}{k}$$

is $o(2^{2^n})$.

Remark. The method used in construction of the lower bound in CPC_1 in Theorem 12 is not applicable to the construction of a bound on CPC_1 for balanced Boolean functions. Therefore the problem of finding better lower bounds on CPC_1 for balanced Boolean functions remains still open.

8 The algebraic degree

The algebraic degree is one of the nonlinearity measures of Boolean function. The Boolean functions with small algebraic degree (linear, quadratic) are in general considered to be less suitable for cryptographic applications than those with higher degree, although there are large classes of cryptographically strong Boolean functions with small algebraic degree (e.g. quadratic bent functions). We prove that almost every (balanced) Boolean function has maximal or almost maximal algebraic degree.

Boolean functions will be represented by their ANF-vectors. Let $\mathrm{tt}(f)$ be the truth table of a Boolean function f, then the corresponding ANF-vector $ANF(f)$ is $\mathrm{tt}(f) \times \mathbf{A}_n$, where $\mathbf{A}_n$ is a binary matrix of order $2^n \times 2^n$ defined recursively:

$$\mathbf{A}_0 = (1); \quad \mathbf{A}_n = \begin{pmatrix} \mathbf{A}_{n-1} & \mathbf{A}_{n-1} \\ \mathbf{0}_{n-1} & \mathbf{A}_{n-1} \end{pmatrix}$$

where $\mathbf{0}_{n-1}$ denotes the zero matrix of order $2^{n-1} \times 2^{n-1}$. Now we can estimate the algebraic degree of random Boolean functions.

Theorem 13. *1. Let f be a random n-ary Boolean function. Then* $\deg(f) \geq n-1$ *almost surely.*

2. Let g be a random n-ary balanced Boolean function. Then $\deg(g) = n-1$ *almost surely.*

Proof. There are $2^{2^n-n-1} = o(2^{2^n})$ Boolean functions with algebraic degree less than $n-1$. Since $2^{2^n-n-1} = o(|\operatorname{Bal}_n|)$ [see Lemma 2], the algebraic degree of almost every n-ary balanced Boolean function is at least $n-1$. Let g be a balanced n-ary Boolean function. The last column in $\mathbf{A}_n$ contains only ones and therefore the last element of its ANF-vector is equal to 0: $a_{2^n-1} = 0$. Consequently, $\deg(g) \neq n$. The theorem follows. □

9 Conclusions

We have shown that almost every n-ary (balanced) Boolean function has such cryptographically strong properties as high nonlinearity, high algebraic degree, correlation immunity and almost optimal propagation characteristics. Since the number of Boolean functions not satisfying a particular criterion (on a sufficiently high level) is $o(2^{2^n})$, we can say that an average n-ary Boolean function is (for a large enough n) cryptographically strong.

Theorem 14. *Let f be an n-ary Boolean function, let g be an n-ary balanced Boolean function and $\phi(n) \to \infty$ as $n \to \infty$. Then the functions f, g have the following properties almost surely*

$$2^{n-1} - 2^{n/2} \cdot \phi(n) \leq \operatorname{wt}(f) \leq 2^{n-1} + 2^{n/2} \cdot \phi(n),$$

$$N_f \geq 2^{n-1} - c_1\sqrt{n} \cdot 2^{n/2};$$

$$N_g \geq 2^{n-1} - c_3\sqrt{n} \cdot 2^{n/2+1};$$

$$CCC_k(f) \geq 2^{n-1} - d_k 2^{n/2}\sqrt{n};$$

$$CCC_k(g) \geq 2^{n-1} - d_k 2^{n/2}\sqrt{n};$$

$$CCC_1(f) \geq 2^{n-1} - c_1 2^{n/2}\sqrt{\lg n};$$

$$CCC_1(g) \geq 2^{n-1} - c_4 2^{n/2+1}\sqrt{\lg n};$$

$$CPC_1(f) \geq 2^{n-1} - 2^{(n+1)/2} \cdot c_1 \cdot \sqrt{\lg(n-1)};$$

$$\deg(f) \geq n-1;$$

$$\deg(g) = n-1.$$

10 Acknowledgement

We would like to thank the anonymous referees whose comments helped in improving the presentation of this paper.

References

[Adams, Tavares 90] Adams C., Tavares S.: "The Structured Design of Cryptographically Good S-Boxes"; Journal of Cryptology, Vol. 3, No. 1, (1990), 27–41.

[Camion, Carlet, Charpin, Sendrier 91] Camion P., Carlet C., Charpin P., Sendrier N.: "On Correlation-immune Functions"; Advances in Cryptology — CRYPTO'91, Springer-Verlag, (1991), 87–100.

[Graham, Knuth, Patashnik 94] Graham R.L., Knuth D.E., Patashnik O.: "Concrete Mathematics: A Foundation for Computer Science"; Addison-Wesley, Second Edition, (1994).

[Millan, Clark, Dawson 97] Millan W., Clark A., Dawson E.: "Smart Hill Climbing Finds Better Boolean Functions"; 4th Workshop on Selected Areas in Cryptography, (1997).

[Mitchell 90] Mitchell Ch.: "Enumerating Boolean functions of Cryptographic Significance"; Journal of Cryptology, Vol. 2, No. 3, (1990), 155–170.

[Pieprzyk, Finkelstein 88] Pieprzyk J., Finkelstein G.: "Towards effective nonlinear cryptosystem design"; IEE Proceedings (Part E), (1988), 135:325–335.

[Siegenthaler 84] Siegenthaler T.: "Correlation-immunity of nonlinear combining functions for cryptographic applications"; IEEE Transactions on Information Theory, IT-30, (1984), 5:776–779.

[Seberry, Zhang, Zheng 93a] Seberry J., Zhang X.-M., Zheng Y.: "Nonlinearity and Propagation Characteristics of Balanced Boolean Functions"; Technical Report no. 4, Computer Security Research Centre, University of Wollongong, Australia, (1993).

[Seberry, Zhang, Zheng 93b] Seberry J., Zhang X.-M., Zheng Y.: "On Constructions and Nonlinearity of Correlation Immune Functions"; Advances in Cryptology – EUROCRYPT'93, Springer-Verlag, (1993), 181–199.

[Tavares 96] Tavares S.E.: personal communication.

[Webster, Tavares 86] Webster A.F., Tavares S.E.: "On the design of S-boxes"; In Advances in Cryptology: Crypto'85 Proceedings, Springer-Verlag, LNCS vol. 219, (1986), 523–534.