On the correlations between a combining function and functions of fewer variables

Anne Canteaut
INRIA - projet CODES,
B.P. 105 - 78153 Le Chesnay Cedex, France.
e-mail: Anne.Canteaut@inria.fr

Abstract — **The Hamming distance of a Boolean function to the functions having many linear structures is an important cryptographic parameter. Most notably, the accuracy of the approximation of the combining function by a function of fewer variables is a major issue in most attacks against combination generators. Here, we show that the distance of a function to the functions having a k-dimensional linear space is highly related to its nonlinearity. In particular, we prove that there is no accurate approximation of any highly nonlinear function by a function depending on a small subset of its input variables.**

I. Introduction

For any Boolean function involved in a symmetric cipher, the existence of linear structures can be exploited in a cryptanalysis. More precisely, the dimension of the *linear space* of a function f, i.e., the dimension of the subspace consisting of all a such that the function $x \mapsto f(x+a)+f(x)$ is constant, has a great cryptanalytic significance. For block ciphers, the complexity of an exhaustive key search can be divided by 2^k when the linear space of the encryption function has dimension k [6]. Filtered registers using a filtering function with linear structures are also vulnerable to some attacks [14]. However, the use of a function without linear structures does not always prevent such attacks. Similar cryptanalytic techniques can still be performed when it is possible to find an accurate approximation of a function by a function having linear structures. Therefore, the Hamming distance of a Boolean function to the set of all functions with linear structures is an important cryptographic criterion, which was investigated by Meier and Staffelbach [9]. Here, we focus on the distance of a function to the set of all functions having a k-dimensional linear space. Indeed, the complexities of most attacks exploiting the existence of linear structures highly depend on the dimension of the linear space.

Amongst all functions having a k-dimensional linear space, the functions for which all linear structures a satisfy $f(x+a)+f(x)=0$, for all x, present an additional weakness. These functions are usually called degenerate because their values only depend on a subset of the input variables. Most notably, the existence of an accurate approximation of a Boolean function by a function depending on fewer variables is a major issue for the cryptanalysis of combination generators.

Combination generators are classical devices for generating a running-key in stream ciphers. They consist of n linear feedback shift registers (LFSR) combined by a nonlinear Boolean function f of n variables. The secret-key of the system is then composed of the initializations of the n constituent LFSRs. It consists of $L_1 + \ldots + L_n$ bits, where L_i denotes the length of the i-th LFSR. Such running-key generators are vulnerable to *correlation attacks* [13]. These "divide-and-conquer" techniques exploit the existence of a correlation between the running-key and the output of one constituent LFSR for recovering the initialization of each LFSR separately. But Siegenthaler's original attack can be prevented by using a *correlation-immune* combining function [12]. In this case, the running-key is statistically independent of the output of each constituent LFSR; any correlation attack should then consider several LFSRs together. More generally, a correlation attack on a set of k LFSRs, namely LFSR i_1, ... , LFSR i_k, exploits the existence of a correlation between the running-key and the output σ of a smaller combination generator, which consists of the k involved LFSRs combined by a Boolean function g of k variables. This attack only succeeds if there exists a function g of k variables whose output is correlated to the output of f, i.e., if there exists g such that

$$p_g = Pr[f(X_1, \ldots, X_n) \neq g(X_{i_1}, \ldots, X_{i_k})] \neq \frac{1}{2}.$$

When such a function exists, different attacks can be performed. In fast correlation attacks [8], the running-key is seen as the result of the transmission of σ through the binary symmetric channel with cross-over probability p_g. Any subsequence of σ is a codeword of a linear code, $\mathcal{C}$, whose dimension corresponds to the linear complexity of σ. In this context, the attack consists in recovering the initializations of the k involved LFSRs by decoding the running-key relatively to $\mathcal{C}$. Therefore, the performance of all techniques for fast correlation attack highly depends on the Hamming distance between f and its best approximation by a function of k variables, g, and on the dimension of the associated code, which is usually equal to $g(L_{i_1}, \ldots, L_{i_k})$ where g is evaluated over integers.

Another technique for recovering the initializations of the k involved LFSRs consists in solving a system of multivariate equations with a Gröbner bases algorithm [7].

Any output bit of the i-th LFSR can obviously be expressed as a linear combination of its initial bits. Therefore, if we approximate f by a function g of k variables, the knowledge of any m running-key bits provides a system of m multivariate equations, which only depends on $L_{i_1} + \ldots + L_{i_k}$ variables (corresponding to the initializations of the k LFSRs $i_1, \ldots, i_k$) and which holds with probability $(1-p_g)^m$. Suppose that the attacker knows $(1-p_g)^m$ different subsets of m running-key bits. Then, the attack consists in successively applying a Gröbner bases algorithm for solving the systems of m equations corresponding to all m-bit subsets of the running-key [4]. Here, the performance of the attack depends on the Hamming distance between the combining function f and its approximation by a function of k variables, g, on the number of involved variables, $L_{i_1}+\ldots+L_{i_k}$, and on the degree of g.

Therefore, the number k of involved registers plays a major role in both attacks: the number of involved variables, $L_{i_1} + \ldots + L_{i_k}$, and the distance of the combining function f to the set of all functions depending on k variables both increase with k. The attacker has then to determine the subset of LFSRs which yields the best trade-off between both parameters. Moreover, these attacks require the approximation g to have a low degree. When the combining function f is t-resilient (i.e., t-th order correlation-immune and balanced), the smallest number of LFSRs that should be considered in the attack is $(t+1)$. In that case, the Boolean function g of $(t+1)$ variables which provides the best approximation of f is the affine function $\sum_{j=1}^{k} x_{i_j} + \varepsilon$ [3, 16]. The number of involved variables and the dimension of the code which appears in fast correlation attacks are then minimal but the error-probability p_g is lower-bounded by $\mathcal{NL}(f)/2^n$ where $\mathcal{NL}(f)$ denotes the nonlinearity of the combining function f. The following natural question then arises: is $(t+1)$ the optimal number of LFSRs that should be considered together in such attacks? The underlying problem is the existence of an accurate approximation of the n-variable combining function f by a function g with 2^{n-k} linear structures a satisfying $g(x+a)+g(x)=0$, for all x. Zhang [16] recently investigated this problem. Most notably, he focused on the approximation of a bent function by a Boolean function of fewer variables. His results point out that the Hamming distance between an n-variable bent function and any k-variable function is always high when the number of involved variables k is small regarding to n. Here, we show that this property holds for any highly nonlinear function, since we prove that the Hamming distance between a function f and the set of all functions having a k-dimensional linear space is highly related to the nonlinearity of f. As a consequence, the optimal number of LFSRs that should be considered together for cryptanalysing a combination generator is $t+1$ when the t-resilient combining function has a high nonlinearity.

II. Preliminaries

We denote by $\mathcal{B}_n$ the set of all Boolean functions of n variables. For any $\alpha \in \mathbf{F}_2^n$, φ_α is the linear function $x \mapsto \alpha \cdot x$ in $\mathcal{B}_n$, where "$\cdot$" denotes the usual dot product. Moreover, $V^\perp$ denotes the dual of a subspace $V \subset \mathbf{F}_2^n$ relatively to this dot product. The Hamming distance between two Boolean functions f and g in $\mathcal{B}_n$ is denoted by $d_H(f,g)$. It corresponds to the number of elements $x \in \mathbf{F}_2^n$ for which $f(x) \neq g(x)$.

A powerful tool for studying the properties of Boolean functions is the *Walsh transform*. We denote by $\mathcal{F}(f)$ the value in 0 of the Walsh transform of $f \in \mathcal{B}_n$:

$$\mathcal{F}(f) = \sum_{x \in \mathbf{F}_2^n} (-1)^{f(x)} = 2^n - 2wt(f)$$

where $wt(f)$ is the Hamming weight of f. A function is said to be *balanced* if $\mathcal{F}(f) = 0$. With the previous notation, the *Walsh spectrum* of a function $f \in \mathcal{B}_n$ is the multiset $\{\mathcal{F}(f+\varphi_\alpha),\ \alpha \in \mathbf{F}_2^n\}$. An important parameter for a Boolean function, which has a great significance for different cryptographic applications, is its nonlinearity.

Definition 1 *The* nonlinearity *of a Boolean function $f \in \mathcal{B}_n$ is its Hamming distance to the set of affine functions. It is equal to can be expressed as*

$$2^{n-1} - \frac{1}{2}\mathcal{L}(f) \text{ where } \mathcal{L}(f) = \max_{\alpha \in \mathbf{F}_2^n} |\mathcal{F}(f+\varphi_\alpha)| \ .$$

Any Boolean function $f \in \mathcal{B}_n$ satisfies $\mathcal{L}(f) \geq 2^{n/2}$; the functions for which equality holds are called *bent functions* [11].

Other cryptographic criteria are related to some properties of the derivatives of a Boolean function.

Definition 2 *Let $f \in \mathcal{B}_n$. For any $a \in \mathbf{F}_2^n$, the* derivative of f with respect to a *is the function $D_a f \in \mathcal{B}_n$ defined by $D_a f(x) = f(x+a)+f(x)$. The* linear space *of f is the subspace of those a such that $D_a f$ is a constant function. Such a nonzero a is called a* linear structure *for f.*

Notation 3 *For any positive integer n and any linear subspace $V \subset \mathbf{F}_2^n$, we denote by $LS(n,V)$ the set of all Boolean functions of n variables whose linear space contains V, i.e.,*

$$LS(n,V) = \{f \in \mathcal{B}_n,\ D_a f \text{ is constant}, \forall a \in V\} \ .$$

Additionally, for any integer $k \leq n$, $LS(n,k)$ is the set of all functions in $\mathcal{B}_n$ whose linear space has dimension at least k. Similarly, we denote by $LS^0(n,V)$ the subset of $LS(n,V)$ defined by

$$LS^0(n,V) = \{f \in \mathcal{B}_n,\ D_a f = 0,\ \forall a \in V\} \ ,$$

and $LS^0(n,k)$ is the set of all functions in $\mathcal{B}_n$ having at least 2^k zero derivatives.

A particular case is the set $\mathcal{B}_n(I)$ of all n-variable functions which only depend on x_i, $i \in I$ where I is a subset of $\{1, \ldots, n\}$. Then, $\mathcal{B}_n(I)$ exactly corresponds

to $LS^0(n,V)$ with $V = Span(e_i, i \in \{1,\ldots,n\} \setminus I)$, where e_i is the i-th canonical basis vector. It follows that the Hamming distance between any function $f \in \mathcal{B}_n$ and the set $\mathcal{B}_n(k)$ of all functions in $\mathcal{B}_n$ depending on k input variables only, $k \leq n$, is lower-bounded by $d_H(f, LS^0(n, n-k))$.

It is worth noticing that the distance of f to the functions of fewer variables may be of little significance in some applications because it is not invariant under the action of the general affine group. As pointed out by Meier and Staffelbach [9], some weaknesses of the involved Boolean function f may be exploited up to equivalence in the sense that some attacks equally apply to f and to any function derived from f by a simple transformation (especially by an affine transformation). It turns out that the quantity that should be considered in such a context is not the distance of f to the set of all functions in $\mathcal{B}_n$ depending on k input variables only, but its distance to the set of all functions in $LS^0(n, n-k)$. Actually, we have

$$d_H(f, LS^0(n,n-k)) = \min_{\pi \in AGL(n)} d_H(f \circ \pi, \mathcal{B}_n(k)) \ ,$$

where $AGL(n)$ is the general affine group.

Moreover, it appears that, for any subspace $V \subset \mathbf{F}_2^n$, the set $LS(n,V)$ can be obtained from $LS^0(n,V)$ by addition of some linear functions:

$$LS(n,V) = \{g + \varphi_b,\ g \in LS^0(n,V), b \in W^\perp\} \ .$$

Therefore, for any function $f \in \mathcal{B}_n$, we have

$$d_H(f, LS(n,V)) = \min_{b \in W^\perp} d_H(f + \varphi_b, LS^0(n,V)) \ .$$

III. Distance of a function to the functions having a k-dimensional linear space

Let f be a Boolean function of n variables. For any subset $V \subset \mathbf{F}_2^n$, we denote by f_V *the restriction of* f *to* V, i.e., the function defined on V by $f_V(x) = f(x)$, for any $x \in V$. When V is a k-dimensional linear subspace of $\mathbf{F}_2^n$, f_V can obviously be identified with a function of k variables. Similarly, for any coset $a+V$ of V, we identify f_{a+V} with $h \in \mathcal{B}_k$ as follows: $h(x) = f(x+a), x \in V$. Then, the *decomposition of* f *with respect to* V is the sequence $\{f_{a+V}\ ,\ a \in W\}$ where $V \times W = \mathbf{F}_2^n$ and all f_{a+V} are considered as Boolean functions in $\mathcal{B}_k$.

Now, we express the distance of a Boolean function to $LS(n,V)$ and to $LS^0(n,V)$ by using its decomposition with respect to V. The following theorem provides a new expression and a generalization of a result due to Zhang [16, Th. 1].

Theorem 4 *Let $f \in \mathcal{B}_n$, let V be a k-dimensional subspace of $\mathbf{F}_2^n$ and let W be such that $W \times V = \mathbf{F}_2^n$. Then, we have*

$$d_H(f, LS^0(n,V)) = 2^{n-1} - \frac{1}{2} \sum_{a \in W} |\mathcal{F}(f_{a+V})| \ .$$

Moreover,

$$d_H(f, LS(n,V)) = 2^{n-1} - \frac{1}{2} \max_{b \in W^\perp} \sum_{a \in W} |\mathcal{F}((f+\varphi_b)_{a+V})|.$$

Therefore, the Hamming distance of f to the set of all functions having a linear space of dimension at least k involves the weights of all elements of the decomposition of f with respect to a k-dimensional subspace. But, these weights are related to Walsh coefficients of f as shown by [2, Th. V.1]. We then deduce a lower bound on the distance of a function f to $LS(n,k)$, which involves the nonlinearity of f.

Theorem 5 *Let $f \in \mathcal{B}_n$ and let V be a k-dimensional subspace of $\mathbf{F}_2^n$. Then,*

$$d_H(f, LS^0(n,V)) \geq 2^{n-1} - \frac{1}{2} \left(\sum_{\alpha \in V^\perp} \mathcal{F}^2(f + \varphi_\alpha) \right)^{\frac{1}{2}} \ .$$

Therefore, the distance of f to the set of all functions having a linear space of dimension at least k satisfies

$$d_H(f, LS(n,k)) \geq 2^{n-1} - 2^{\frac{n-k}{2}-1} \mathcal{L}(f)$$

where $\mathcal{L}(f)$ is the highest magnitude occurring in the Walsh spectrum of f, i.e., $\mathcal{L}(f) = \max_{\alpha \in \mathbf{F}_2^n} |\mathcal{F}(f + \varphi_\alpha)|$.

The previous theorem clearly points out that any highly nonlinear function lies at high Hamming distance to the set of all functions having a k-dimensional linear space, when k is large. In particular, *when a function f has a high nonlinearity, there is no accurate approximation of f by a function depending on a small subset of its input variables.* Note that, for bent functions, the previous theorem directly improves Zhang's result [16, Theorem 5].

IV. Approximations of resilient functions and of functions satisfying the propagation criterion

When f is used in a combination generator, the distribution probability of its output should be unaltered when any t of its inputs are fixed [13]. This property is called *t-th order correlation-immunity* [12]. Balanced t-th order correlation-immune functions are called *t-resilient* functions. Such functions are characterized by the set of zero values in their Walsh spectra [15]. Now, Theorem 5 allows to estimate the distance of a resilient function to all functions depending on fewer variables.

Theorem 6 *Let $f \in \mathcal{B}_n$ be a t-resilient function. Then, the Hamming distance of f to the set $\mathcal{B}_n(k)$ of all functions depending on k input variables satisfies:*

$$d_H(f, \mathcal{B}_n(k)) \geq 2^{n-1} - \frac{\mathcal{L}(f)}{2} \left(\sum_{i=t+1}^{k} \binom{k}{i} \right)^{\frac{1}{2}} \ .$$

Note that this theorem clearly implies that any t-resilient function is uncorrelated to any function depending on t variables only. Moreover, we also recover from Theorem 5 that the Hamming distance of a t-resilient function to the functions depending on $(t+1)$ variables, $x_i, i \in I$,

is equal to $2^{n-1} - \frac{1}{2}|\mathcal{F}(f+\varphi_u)|$ where u is the n-bit vector of support I [3, 16]. Theorem 5 actually proves that this value is a lower bound for $d_H(f, \mathcal{B}(I))$, and it is an upper bound too since it is achieved by the affine function $\sum_{i\in I} x_i + \varepsilon$.

Some applications require that the output difference of a Boolean function be uniformly distributed for low-weight input differences. This property, referred as *propagation criterion* [10], is notably important when the function is used in a hash function or in a block cipher. This criterion involves some properties of the derivatives of f.

Definition 7 *A function $f \in \mathcal{B}_n$ satisfies the* propagation criterion of degree t (PC(t)) *if $\mathcal{F}(D_\alpha f) = 0$ for all $\alpha \in \mathbf{F}_2^n$ such that $1 \leq wt(\alpha) \leq t$.*

The propagation criterion is clearly related to the distance to the functions having linear structures. This distance is maximal if and only if the involved function satisfies $PC(n)$ [9]. More generally, when f satisfies the propagation criterion of degree t, the lower bound on its distance to the set of all functions depending on k variables can be improved, especially for $k \geq n-t$, by using [1, Prop. 5].

Theorem 8 *Let $f \in \mathcal{B}_n$ be a function satisfying $PC(t)$. Then, the Hamming distance of f to the set $\mathcal{B}_n(k)$ of all functions depending on k input variables satisfies:*

$$d_H(f, \mathcal{B}_n(k)) \geq 2^{n-1} - 2^{\frac{k}{2}-1}\left(2^n + \mathcal{M}(f)\sum_{i=t+1}^{n-k}\binom{k}{i}\right)^{\frac{1}{2}}$$

where $\mathcal{M}(f) = \max_{e\neq 0}|\mathcal{F}(D_e f)|$.

Most notably, if $k \geq n-t$, we have

$$d_H(f, \mathcal{B}_n(k)) \geq 2^{n-1} - 2^{\frac{n+k}{2}-1} .$$

Note that this theorem provides the same upper bound as Theorem 5 for bent functions, since n-variable bent functions satisfy $PC(n)$.

V. Distance of a bent function to the functions having a k-dimensional linear space

The case of bent functions is of most interest because they lie as far as possible from the functions having linear structures. Moreover, it clearly appears from Theorem 5 that their distance to the functions having a k-dimensional subspace is high, because it is lower-bounded by $2^{n-1} - 2^{n-1-\frac{k}{2}}$. However, bent functions may have different behaviors in the sense that the accuracy of their best approximation by a function having 2^k linear structures may vary when $k > 1$. Now, we characterize the bent functions which lie as close as possible to the functions having a k-dimensional linear space, i.e., the functions which achieve the previous lower bound. This characterization is related to some properties of the *dual function*.

Definition 9 *[5] Let f be a bent function in $\mathcal{B}_n$. Then, there exists a Boolean function $\widetilde{f} \in \mathcal{B}_n$, called the* dual *of f, such that*

$$\mathcal{F}(f+\varphi_\alpha) = 2^{\frac{n}{2}}(-1)^{\widetilde{f}(\alpha)} \text{ for all } \alpha \in \mathbf{F}_2^n .$$

Theorem 10 *Let $f \in \mathcal{B}_n$ be a bent function and let V be a k-dimensional subspace of $\mathbf{F}_2^n$. Then,*

$$d_H(f, LS(n, V)) \geq 2^{n-1} - 2^{n-1-\frac{k}{2}}$$

where equality holds if and only if k is even and the restriction of the dual function, $\widetilde{f}$, to a coset of $V^\perp$ is bent.

References

[1] A. Canteaut, C. Carlet, P. Charpin, and C. Fontaine. Propagation characteristics and correlation-immunity of highly nonlinear Boolean functions. In *Advances in Cryptology - EUROCRYPT 2000*, LNCS 1807, pages 507–522. Springer-Verlag, 2000.

[2] A. Canteaut, C. Carlet, P. Charpin, and C. Fontaine. On cryptographic properties of the cosets of $R(1, m)$. *IEEE Trans. Inform. Theory*, 47(4):1494–1513, 2001.

[3] A. Canteaut and M. Trabbia. Improved fast correlation attacks using parity-check equations of weight 4 and 5. In *Advances in Cryptology - EUROCRYPT 2000*, LNCS 1807, pages 573–588. Springer-Verlag, 2000.

[4] N.T. Courtois. Higher order correlation attacks, XL algorithm and cryptanalysis of Toyocrypt. Preprint, 2002. Available at `http://eprint.iacr.org/2002/087`.

[5] J.F. Dillon. *Elementary Hadamard Difference sets*. PhD thesis, University of Maryland, 1974.

[6] J.H. Evertse. Linear structures in block ciphers. In *Advances in Cryptology - EUROCRYPT'87*, LNCS 304, pages 249–266. Springer-Verlag, 1988.

[7] J.-C. Faugère. A new efficient algorithm for computing Gröbner bases without reduction to zero (F5). In *International Symposium on Symbolic and Algebraic Computation (ISSAC 2002)*, 2002.

[8] W. Meier and O. Staffelbach. Fast correlation attack on certain stream ciphers. *J. Cryptology*, pages 159–176, 1989.

[9] W. Meier and O. Staffelbach. Nonlinearity criteria for cryptographic functions. In *Advances in Cryptology - EUROCRYPT'89*, LNCS 434, pages 549–562. Springer-Verlag, 1990.

[10] B. Preneel, W.V. Leekwijck, L.V. Linden, R. Govaerts, and J. Vandewalle. Propagation characteristics of Boolean functions. In *Advances in Cryptology - EUROCRYPT'90*, LNCS 437, pages 155–165. Springer-Verlag, 1990.

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

[12] T. Siegenthaler. Correlation-immunity of nonlinear combining functions for cryptographic applications. *IEEE Trans. Inform. Theory*, IT-30(5):776–780, 1984.

[13] T. Siegenthaler. Decrypting a class of stream ciphers using ciphertext only. *IEEE Trans. Computers*, C-34(1):81–84, 1985.

[14] T. Siegenthaler. Cryptanalysts representation of nonlinearly filtered ML-sequences. In *Advances in Cryptology - EUROCRYPT'85*, LNCS 219, pages 103–110. Springer-Verlag, 1986.

[15] G. Xiao and J.L. Massey. A spectral characterization of correlation-immune combining functions. *IEEE Trans.Inform. Theory*, IT-34(3):569–571, 1988.

[16] M. Zhang. Maximum correlation analysis of nonlinear combining functions in stream ciphers. *Journal of Cryptology*, 13(3):301–313, 2000.