Four Measures of Nonlinearity[⋆]

Joan Boyar[1⋆⋆], Magnus Find[1⋆⋆⋆], and René Peralta[2]

[1] Department of Mathematics and Computer Science,
University of Southern Denmark, Denmark
{joan,magnusgf}@imada.sdu.dk
[2] Information Technology Laboratory,
National Institute of Standards and Technology, USA
peralta@nist.gov

Abstract. Cryptographic applications, such as hashing, block ciphers and stream ciphers, make use of functions which are simple by some criteria (such as circuit implementations), yet hard to invert almost everywhere. A necessary condition for the latter property is to be "sufficiently distant" from linear, and cryptographers have proposed several measures for this distance. In this paper, we show that four common measures, *nonlinearity, algebraic degree, annihilator immunity*, and *multiplicative complexity*, are incomparable in the sense that for each pair of measures, μ_1, μ_2, there exist functions f_1, f_2 with $\mu_1(f_1) > \mu_1(f_2)$ but $\mu_2(f_1) < \mu_2(f_2)$. We also present new connections between two of these measures. Additionally, we give a lower bound on the multiplicative complexity of collision-free functions.

1 Preliminaries

For a vector $\mathbf{x} \in \mathbb{F}_2^n$ its *Hamming weight* is the number of non-zero entries in $\mathbf{x}$. For $n \in \mathbb{N}$ its Hamming weight, $H^{\mathbb{N}}(n)$ is defined as the Hamming weight of the binary representation of n. We let $B_n = \{f : \mathbb{F}_2^n \to \mathbb{F}_2\}$ be the set of Boolean predicates on n variables.

A Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ can be uniquely represented by its *algebraic normal form* also known as its Zhegalkin polynomial [30]:

$$f(x_1, \ldots, x_n) = \bigoplus_{S \subseteq \{1,2,\ldots,n\}} \alpha_S \prod_{i \in S} x_i$$

where $\alpha_s \in \{0, 1\}$ for all S and we define $\prod_{i \in \emptyset} x_i$ to be 1. If $\alpha_S = 0$ for $|S| > 1$, we say that f is *affine*. An affine function f is *linear* if $\alpha_\emptyset = 0$ or equivalently if $f(\mathbf{0}) = 0$. The function f is *symmetric* if $\alpha_S = \alpha_{S'}$ whenever $|S| = |S'|$, that

[⋆] This version corrects some typos in the CIAC 2013 version of the paper. It also contains the proofs missing in that version.

[⋆⋆] Partially supported by the Danish Council for Independent Research, Natural Sciences. Part of this work was done while visiting the University of Waterloo.

[⋆⋆⋆] Part of this work was done while visiting the University of Toronto.

is f only depends on the Hamming weight of the input. The kth *elementary symmetric Boolean function*, denoted Σ_k^n, is defined as the sum of all terms where $|S| = k$.

For two functions $f, g \in B_n$ the distance d between f and g is defined as the number of inputs where the functions differ, that is

$$d(f,g) = |\{\mathbf{x} \in \mathbb{F}_2^n | f(\mathbf{x}) \neq g(\mathbf{x})\}| \,.$$

For the rest of this paper, unless otherwise stated, n denotes the number of input variables. We let log denote the logarithm base 2 and ln the natural logarithm.

2 Introduction

Cryptographic applications, such as hashing, block ciphers and stream ciphers, make use of functions which are simple by some criteria (such as circuit implementations) yet hard to invert almost everywhere. A necessary condition for the latter to hold is that the tools of algebra – and in particular linear algebra – be somehow not applicable to the problem of saying something about x given $f(x)$. Towards this goal, cryptographers have proposed several measures for the distance to linearity for Boolean functions. In this paper we consider four such measures. We compare and contrast them, both in general and in relation to specific Boolean functions. Additionally, we propose a procedure to find collisions when the multiplicative complexity is low.

The *nonlinearity* of a function is the Hamming distance to the closest affine function. The nonlinearity of a function on n bits is between 0 and $2^{n-1} - \lceil 2^{n/2-1} \rceil$ [26, 6]. Affine functions have nonlinearity 0. Unfortunately, this introduces an overloading of the word "nonlinearity" since it also refers to the more general concept of distance to linear. The meaning will be clear from context.

Functions with nonlinearity $2^{n-1} - 2^{n/2-1}$ exist if and only if n is even. These functions are called *bent*, and several constructions for bent functions exist (see [26, 21, 14] or the survey by Carlet [6]). For odd n, the situation is a bit more complicated; for any bent function f on $n-1$ variables, the function $g(x_1, \ldots, x_n) = f(x_1, \ldots, x_{n-1})$ will have nonlinearity $2^{n-1} - 2^{(n-1)/2}$. It is known that for odd $n \geq 9$, this is suboptimal [17]. Despite this, no infinite family achieving higher nonlinearity is known. For a Boolean function f, there is a tight connection between the nonlinearity of f and its Fourier coefficients. More precisely the nonlinearity is determined by the largest Fourier coefficient, and for bent functions all the Fourier coefficients have the same magnitude. A general treatment on Fourier analysis, can be found in [24].

The *algebraic degree* (which we from now on will refer to as just the *degree*) of a function is the degree of its Zhegalkin polynomial, that is the largest $|S|$ such that $\alpha_S = 1$. We note that Carlet [5] has compared nonlinearity and degree to two other measures which we do not consider here, algebraic thickness and nonnormality.

The *annihilator immunity* (also known as algebraic immunity[3]) of a function f is the minimum degree of a non-zero function g such that $fg = 0$ or $(f+1)g = 0$. We denote this measure by $AI(f)$. The function g is called an *annihilator*. It is known that $0 \leq AI(f) \leq \lceil \frac{n}{2} \rceil$ for all functions [8, 10]. Specific functions are known which achieve the upper bound [11].

The *multiplicative complexity* of a function f, denoted $c_\wedge(f)$, is the smallest number of AND gates necessary and sufficient to compute the function using a circuit over the basis (XOR,AND,1) (i.e. using arithmetic over $GF(2)$). Clearly, the multiplicative complexity of f is at least 0 with equality if and only if f is affine. For even n, the multiplicative complexity is at most $2^{\frac{n}{2}+1} - \frac{n}{2} - 2$, and for odd n at most $\frac{3}{2\sqrt{2}} 2^{n/2+1} - \frac{n+3}{2}$ [3, 22] (see also [16]). Despite this, no specific predicate has been proven to have multiplicative complexity larger than $n-1$.[4]

Nonlinearity, degree and multiplicative complexity all capture an intuitive notion of the degree of "nonlinearity" of Boolean functions. Annihilator immunity is also related to nonlinearity, albeit less obviously.

In [6], it is shown that algebraic degree, annihilator immunity, and nonlinearity are affine invariants. That is, if $L : \{0,1\}^n \to \{0,1\}^n$ is an invertible linear mapping, applying L to the input variables first does not change the value of any of these measures. It is easy to see that multiplicative complexity is also an affine invariant, since L and L^{-1} can be computed using only XOR gates.

Ideally, a measure of nonlinearity should be invariant with respect to addition of affine functions and embedding into a higher dimensional space (e.g. considering $f(x_1, x_2) = x_1 x_2$ as a function of three variables). The four measures studied here have these properties with two exceptions.

- Adding an affine function l to f can cause the annihilator immunity to vary by up to 1. That is $AI(f) - 1 \leq AI(f + l) \leq AI(f) + 1$ [7];
- Embedding a function $f : \{0,1\}^n \to \{0,1\}$ in $\{0,1\}^{n+1}$ doubles its nonlinearity. Thus, if one wants to consider nonlinearity of functions embedded in larger spaces, it might be more natural to redefine nonlinearity using a normalized metric instead of the Hamming distance metric. In this paper, we will not use embeddings.

There is a substantial body of knowledge which relates nonlinearity, annihilator immunity, and algebraic degree to cryptographic properties. However, the analogous question with respect to multiplicative complexity remains little studied. Among the few published results is [9], in which Courtois et al. show (heuristically) that functions with low multiplicative complexity are less resistant

[3] In this paper we use the term "annihilator immunity" rather than "algebraic immunity", see the remark in [11].

[4] We have experimentally verified that all predicates on four bits have multiplicative complexity at most three. This is somewhat surprising, as circuit realization of random functions (e.g. $x_1x_2x_3x_4 + x_1x_2x_3 + x_2x_3x_4 + x_1x_3x_4 + x_1x_3 + x_2x_4 + x_1x_4$) would appear to need more than three AND gates. We conjecture that some predicate on five bits will turn out to have multiplicative complexity five.

against algebraic attacks. Here we present evidence that low multiplicative complexity in hash functions can make them prone to second preimage or collision attacks.

Multiplicative complexity also turns out to be important in cryptographic protocols. Several techniques for secure multi-party computation yield protocols with communication complexity proportional to the multiplicative complexity of the function being evaluated (see, for example, [15, 18, 23]). Several flavors of one-prover non-interactive cryptographically secure proofs (of knowledge of x given $f(x)$) have length proportional to the multiplicative complexity of the underlying function f (see, for example, [1]).

In this paper we show that very low nonlinearity implies low multiplicative complexity and vice-versa. We also show an upper bound on nonlinearity for functions with very low multiplicative complexity.

For nonlinearity, annihilator immunity, and algebraic degree, there exist symmetric Boolean functions achieving the maximal value among all Boolean functions. However, the only symmetric functions which achieve maximum nonlinearity are the quadratic functions, which have low algebraic degree. In [4] Canteaut and Videau have characterized the symmetric functions with almost optimal nonlinearity. In this paper we analyze the multiplicative complexity and annihilator immunity of these functions.

3 Relations between Nonlinearity Measures

In general, random Boolean functions are highly nonlinear with respect to all these measures:

- In [13], Didier shows that the annihilator immunity of almost every Boolean function is $(1 - o(1))n/2$.
- In [25], Rodier shows that the nonlinearity of almost every function is at least $2^{n-1} - 2^{n/2-1}\sqrt{2n \ln 2}$, which is close to maximum.
- In [5], Carlet observes that almost every function has degree at least $n - 1$.
- In [3], Boyar et al. show that almost every Boolean function has multiplicative complexity at least $2^{n/2} - O(n)$.

If a function f has algebraic degree d, the multiplicative complexity is at least $d - 1$ [28]. This is a very weak bound for most functions. However this technique easily yields lower bounds of $n - 1$ for many functions on n variables, and no larger lower bounds are known for concrete functions

Additionally, it has been shown that low nonlinearity implies low annihilator immunity [10]. Still, there are functions optimal with respect to annihilator immunity that have nonlinearity much worse than that of bent functions. An example of this is the majority function, see [11]. Bent functions have degree at most $\frac{n}{2}$ ([26, 6]). Since $f \oplus 1$ is an annihilator for f, the annihilator immunity of a function is at most its degree.

4 Incomparability

In this section we show that our four measures are incomparable in the sense that for each pair of measures, μ_1, μ_2, there exist functions f_1, f_2 with $\mu_1(f_1) > \mu_1(f_2)$, but $\mu_2(f_1) < \mu_2(f_2)$. To show this we look at four functions:

Σ_2^n: For even n, the function Σ_2^n is bent [26]. For odd n it has nonlinearity $2^{n-1} - 2^{(n-1)/2}$, which is maximum among the symmetric functions on an odd number of variables [20]. But being a quadratic function, both the algebraic degree and the annihilator immunity are 2 which is almost as bad as for linear functions. The multiplicative complexity is $\lfloor n/2 \rfloor$, which is the smallest possible multiplicative complexity for nonlinear symmetric functions [3].

MAJ_n, which is 1 if and only if at least $n/2$ of the n inputs are 1: In [2] it is shown that when $n = 2^r + 1$, the multiplicative complexity is at least $n - 2$. In [11] it is shown that MAJ_n has annihilator immunity $\lceil \frac{n}{2} \rceil$; they also show that it has nonlinearity $2^{n-1} - \binom{n-1}{\lfloor \frac{n}{2} \rfloor}$, which by Stirling's approximation is $2^{n-1} - (1 + o(1))\sqrt{\frac{2}{\pi}} \frac{2^{n-1}}{\sqrt{n-1}}$.

$FMAJ_n$, defined as:

$$FMAJ_n(x_1, \ldots, x_n) = MAJ_{\lceil \log n \rceil}(x_1, \ldots, x_{\lceil \log n \rceil}) \oplus x_{\lceil \log n \rceil + 1} \oplus \ldots \oplus x_n.$$

The degree of $FMAJ_n$ is equal to the degree of $MAJ_{\lceil \log n \rceil}$ which is at least $\frac{\lceil \log n \rceil}{2}$, so the multiplicative complexity is at least $\frac{\lceil \log n \rceil}{2} - 1$. Also its multiplicative complexity is equal to that of $MAJ_{\lceil \log n \rceil}$, which is at most $\lceil \log(n) \rceil - H^{\mathbb{N}}(\lceil \log n \rceil) + \lceil \log(\lceil \log n \rceil + 1) \rceil$ [2]. The annihilator immunity of $FMAJ_n$ is at least $\left\lceil \frac{\log n}{2} \right\rceil - 1$, since $MAJ_{\lceil \log n \rceil}$ has annihilator immunity $\left\lceil \frac{\log n}{2} \right\rceil$, and $FMAJ_n$ is just $MAJ_{\lceil \log n \rceil}$ plus a linear function. This can change the annihilator immunity by at most 1 [7].

Σ_n^n : The nonlinearity of Σ_n^n is 1 because it has Hamming distance 1 to the zero function. It has annihilator immunity 1 ($x_1 \oplus 1$ is an annihilator), its algebraic degree is n, and its multiplicative complexity is $n - 1$.

Incomparability examples: From the observations above it can be seen that Σ_2^n has higher nonlinearity than MAJ_n but smaller degree, annihilator immunity, and multiplicative complexity. $FMAJ_n$ has higher degree and annihilator immunity than Σ_2^n but lower multiplicative complexity. Σ_n^n has larger degree than $FMAJ_n$ but smaller annihilator immunity. These examples are shown in Table 1. *Remark:* These separations are fairly extreme except with respect to multiplicative complexity, where the values are small compared to those for random functions. This is due to the fact that currently no specific function has been proven to have multiplicative complexity larger than $n - 1$. If larger bounds were proven, one could have more extreme separations: Suppose $f : \{0,1\}^{n-1} \to \{0,1\}$ has large multiplicative complexity, degree, nonlinearity and annihilator immunity, and let $g(x_1, \ldots, x_n) = f(x_1, \ldots, x_{n-1}) \cdot x_n$. Then clearly g has high degree, nonlinearity and multiplicative complexity, but annihilator immunity 1, since multiplying by $x_n + 1$ gives the zero function. This is also an example where the annihilator immunity fails to capture the intuitive notion of nonlinearity.

Table 1. Incomparability examples. For every pair (f_1, f_2) f_1 scores higher in the measure for the row and f_2 scores higher in the measure for the column.

	NL	MC	deg	AI
NL	-	(Σ_2^n, MAJ_n)	(Σ_2^n, MAJ_n)	(Σ_2^n, MAJ_n)
MC	-	-	$(\Sigma_2^n, FMAJ_n)$	$(\Sigma_2^n, FMAJ_n)$
deg	-	-	-	$(\Sigma_n^n, FMAJ_n)$

5 Relationship between Nonlinearity and Multiplicative Complexity

In this section we will show that, despite being incomparable measures, the multiplicative complexity and nonlinearity are somehow related. We first show that if a function has low nonlinearity, this gives a bound on the multiplicative complexity. Conversely we show that if a function $f \in B_n$ has multiplicative complexity $M \leq \frac{n}{2}$, it has nonlinearity at most $2^{n-1} - 2^{n-M-1}$. Furthermore for $M \leq \frac{n}{2}$, there exist a simple function with this nonlinearity.

We will use the following theorem due to Lupanov [19] (see Lemma 1.2 in [16]). Given a Boolean matrix A, a *decomposition* is a set of Boolean matrices $B_1, \ldots, B_k$ each having rank 1, satisfying $A = B_1 + B_2 + \ldots + B_k$ where addition is over the reals. For each B_i its weight is defined as the number of non-zero rows plus the number of non-zero columns. The *weight* of a decomposition is the sum of the weights of the B_i's.

Theorem 1 (Lupanov). *Every Boolean $p \times q$ matrix admits a decomposition of weight*

$$(1 + o(1))\frac{pq}{\log p}.$$

Theorem 2. *A function $f \in B_n$ with nonlinearity $s > 1$ has multiplicative complexity at most* $\min\{s(n-1), (2 + o(1))\frac{sn}{\log s}\}$.

Proof. Let L be an affine function with minimum distance to f. Let

$$\epsilon(\mathbf{x}) = f(\mathbf{x}) \oplus L(\mathbf{x}).$$

Note that ϵ takes the value 1 s times. Let $\epsilon^{-1}(1)$ be the preimage of 1 under ϵ. Suppose $\epsilon^{-1}(1) = \{z^{(1)}, \ldots, z^{(s)}\}$ where each $z^{(i)}$ is an n-bit vector. Let $M_i(\mathbf{x}) = \prod_{j=1}^{n}(x_j \oplus z_j^{(i)} \oplus 1)$ be the minterm associated to $z^{(i)}$, that is the polynomial that is 1 only on $z^{(i)}$. By definition

$$\epsilon(\mathbf{x}) = \bigoplus_{i=1}^{s} M_i(\mathbf{x}) = \bigoplus_{i=1}^{s}\prod_{j=1}^{n}(x_j \oplus z_j^{(i)} \oplus 1)$$

Adding the minterms together can be done using only XOR gates and gives exactly the function ϵ. We will give two constructions for the minterms. Using the one with fewest AND gates proves the result.

The first construction simply computes each of the s minterms directly using $n-1$ AND gates for each. For the second construction, define the $s \times 2n$ matrix A where columns $1, 2, \ldots, n$ correspond to $x_1, x_2, \ldots, x_n$ and columns $n+1, \ldots, 2n$ correspond to $(1 \oplus x_1), \ldots, (1 \oplus x_n)$, and row i corresponds to minterm M_i. Let $A_{ij} = 1$ if and only if the literal corresponding to column j is a factor in the minterm M_i. Now consider the rectangular decomposition guaranteed to exist by Theorem 1. For each B_i, all non-zero columns are equal. AND together the literals corresponding to these variables. Call the result Q_i. Now each row can be seen as a logical AND of Q_i's. AND these together for every row to obtain the s results. The number of AND gates used is at most the weight of the decomposition, that is at most $(1+o(1))\frac{2sn}{\log s}$ AND gates. □

Lemma 1. *Let f have multiplicative complexity $M \leq \frac{n}{2}$. Then there exists an invertible linear mapping $L : \{0,1\}^n \to \{0,1\}^n$, a Boolean predicate $g \in B_D$ for $D \leq 2M$, and a set $T \subseteq \{1, 2, \ldots, n\}$ such that for $\mathbf{t} = L(\mathbf{x})$, f can be written as*

$$f(x_1, \ldots, x_n) = g(t_1, \ldots, t_D) \oplus \bigoplus_{j \in T} t_j$$

Proof. Let $M = c_\wedge(f)$ and consider an XOR-AND circuit C with M AND gates computing f, and let $A_1, \ldots, A_M$ be a topological ordering of the AND gates. Let the inputs to A_1 be I_1, I_2 and inputs to A_2 be I_3, I_4, etc. so A_M has inputs I_{2M-1}, I_{2M}. Now the value of f, the output of C, can be written as a sum of some of the AND gate outputs and some of the inputs to the circuit:

$$f = \bigoplus_{i \in Z_{out}} A_i \oplus \bigoplus_{i \in X_{out}} x_i,$$

for appropriate choices of Z_{out} and X_{out}. Similarly for I_j:

$$I_j = \bigoplus_{i \in Z_j} A_i \oplus \bigoplus_{i \in X_j} x_i.$$

Define g as $g = \bigoplus_{i \in Z_{out}} A_i$. Since X_j is a subset of $\{0,1\}^n$, it can be thought of as a vector y_j in the vector space $\{0,1\}^n$ where the ith coordinate is 1 if and only if $i \in X_j$.

Clearly the dimension D of $Y = \mathit{span}(y_1, \ldots y_{2M})$ is at most $2M$. Let $\{y_{j_1}, \ldots y_{j_D}\}$ be a basis of Y. There exists some invertible linear mapping $L : \{0,1\}^n \to \{0,1\}^n$ with $L(x_1, \ldots, x_n) = (t_1, \ldots, t_n)$ having $t_j = y_{i_j}$ for $1 \leq j \leq D$. That is, g depends on just $t_1, \ldots t_D$, and each x_j is a sum of t_l's, hence f can be written as a function of $t_1, \ldots, t_n$ as

$$f = g(t_1, \ldots, t_D) \oplus \bigoplus_{j \in T} t_j$$

□

Corollary 1. *If a function $f \in B_n$ has multiplicative complexity $M \leq \frac{n}{2}$, it has nonlinearity at most $2^{n-1} - 2^{n-M-1}$. Furthermore for $M \leq \frac{n}{2}$, there exist a simple function with this nonlinearity.*

Proof. Since nonlinearity is an affine invariant, we can use Lemma 1 and look at the nonlinearity of

$$f = g(t_1, \ldots, t_{2M}) \oplus \bigoplus_{j \in T_{out}} t_j$$

Now the best affine approximation of g agrees on at least $2^{2M-1} + 2^{M-1}$ inputs. Replacing g with its best affine approximation, we obtain a function that agrees with f on at least $2^{n-2M}(2^{2M-1} + 2^{M-1}) = 2^{n-2M}2^{2M-1} + 2^{n-2M}2^{M-1} = 2^{n-1} + 2^{n-M-1}$ proving the upper bound on the nonlinearity. For the furthermore part notice that the nonlinearity of the function

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{M} x_{2i-1}x_{2i}$$

meets the bound. □

Remark: This shows that Σ_2^n is optimal with respect to nonlinearity among functions having multiplicative complexity $\lfloor n/2 \rfloor$.

6 Low Multiplicative Complexity and One-Wayness

If a function f has multiplicative complexity μ, then it can be inverted (i.e. a preimage can be found) in at most 2^μ evaluations of f. To do this, consider a circuit C for f with μ AND gates. Suppose y has a non-empty preimage under f. Guessing the Boolean value of one input for each AND gate results in a linear system of equations, L. Solve L to obtain a candidate input x and test whether $f(x) = y$. This finds a preimage of y after at most 2^μ iterations. Thus, one-way functions, if they exist, have superlogarithmic multiplicative complexity.

The one-wayness requirements of hash functions include the much stronger requirement of *collision resistance*: it must be infeasible to find two inputs that map to the same output. We next observe that collision resistance of a function f with n inputs and $m < n$ outputs requires f to have multiplicative complexity at least $n - m$.

Let C be a circuit for f. Without loss of generality, we can assume the circuit contains no negations and that we seek two distinct inputs which map to $\mathbf{0}$. [5] Since there are no negations in the circuit, one such input is $\mathbf{0}$. We next show how to obtain a second preimage of $\mathbf{0}$.

Pick a topologically minimal AND gate and set one of its inputs to 0. This generates one homogeneous linear equation on the inputs to f and allows us to remove the AND gate from the circuit (see Figure 1). Repeating this until no AND gates are left yields a homogeneous system S with at most μ equations,

[5] Negations can be “pushed” to the outputs of the circuit without changing the number of AND gates. Once at the outputs, for purposes of finding a collision, negations can be simply removed.

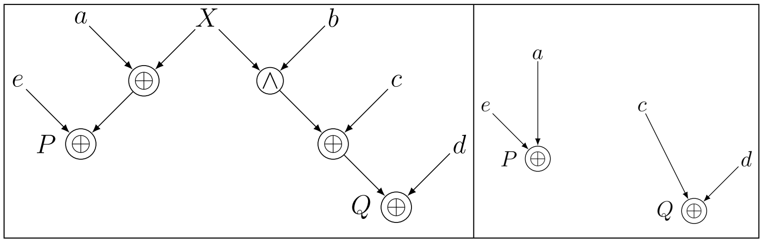

Fig. 1. The circuit to the right is the circuit obtained when X in the left circuits is restricted to the value 0. Notice that only the gates P, Q remain nonredundant.

plus a circuit C' which computes a homogeneous linear system with m equations. The system of equations has $2^{n-m-\mu}$ distinct solutions. Thus, if $m+\mu < n$, then standard linear algebra yields non-zero solutions. These are second preimages of $\mathbf{0}$.

We re-state this as a theorem below. The idea of using hyperplane restrictions to eliminate AND gates has been used before, however with different purposes, see e.g. [2, 12].

Theorem 3. *Collision resistance of a function f from n to m bits requires that f have multiplicative complexity at least $n - m$.*

It is worth noting that the bound from Theorem 3 does not take into account the position of the AND gates in the circuit. It is possible that fewer linear equations can be used to remove all AND gates. We have tried this on the reduced-round challenges issued by the Keccak designers (Keccak is the winner of the SHA-3 competition, see `http://keccak.noekeon.org/crunchy_contest.html`). These challenges are described in the notation Keccak$[r, c, nr]$ where r is the rate, c the capacity, and nr the number of rounds. For the collision challenges, the number of outputs is set to 160. Each round of Keccak uses $r+c$ AND gates. However, in the last round of Keccak the number of AND gates that affect the output bits is equal to the number of outputs.

We consider circuits for Keccak with only one block (r bits) of input. The circuit for Keccak$[r{=}1440, c{=}160, nr{=}1]$ contains 160 AND gates, yet 96 linear equations will remove them all. Keccak$[r{=}1440, c{=}160, nr{=}2]$ contains 1760 AND gates, yet 1056 linear equations removes them all. Thus, finding collisions is easy, because 1440 is greater than 160 + 1056 (in the one-round case, because 1440 > 160 + 96). These two collision challenges were first solved by Morawiecki (using SAT solvers, see `http://keccak.noekeon.org/crunchy_mails/coll-r2-w1600-20110729.txt`) and, more recently, by Duc et al. (see `http://keccak.noekeon.org/crunchy_mails/coll-r1r2-w1600-20110802.txt`). Our reduction technique easily solves both of these challenges, and yields a large number of multicollisions.

Dinur et al. are able to obtain collisions for Keccak$[r{=}1440, c{=}160, nr{=}4]$ i.e. for four rounds of Keccak (see `http://keccak.noekeon.org/crunchy_mails/coll-r3r4-w1600-20111124.txt`). The technique of Theorem 3 cannot linearize the Keccak circuit for more than two rounds. How to leverage our methods to solve three or more rounds is work in progress.

7 Some Symmetric Boolean Functions with High Nonlinearity

When designing Boolean functions for cryptographic applications, we seek functions with high nonlinearity, simple structure, high annihilator immunity, and high algebraic degree. Bent functions have high nonlinearity. Symmetric functions have simple structure. However, the multiplicative complexity of a symmetric function on n variables is never larger than $n + 3\sqrt{n}$ [3]. The symmetric functions with highest nonlinearity are quadratic ([27] and [20]). But these functions have low algebraic degree, low annihilator immunity, and multiplicative complexity only $\lfloor \frac{n}{2} \rfloor$.

For $n \geq 3$, let $F_n = \bigoplus_{k=3}^{n} \Sigma_k^n$ and $G_n = \Sigma_2^n \oplus \Sigma_n^n$. It is known that there are exactly 8 symmetric functions with nonlinearity exactly 1 less than the largest achievable value. These are $F_n \oplus \lambda$ and $G_n \oplus \lambda$, where $\lambda \in \{0, 1, \Sigma_1^n, \Sigma_1^n + 1\}$ [4]. These functions have many of the criteria sought after for cryptographic functions: they are symmetric, have optimal degree, and almost optimal nonlinearity. We have exactly calculated or tightly bound the multiplicative complexity of these functions. Precise values are important for applications in secure multiparty computations.

Since the λ can always be computed and added using only XOR operations, we only consider F_n and G_n. In [2] it is shown that the Hamming weight of n bits $x_1, \ldots, x_n$ can be computed using an XOR-AND circuit having $n - H^{\mathbb{N}}(n)$ AND gates, where $H^{\mathbb{N}}(n)$ is the Hamming weight of the binary representation of n. Furthermore, it is noted that the value of the ith least significant bit in the Hamming weight is equal to the function $\Sigma_{2^i}^n(x_1, \ldots, x_n)$ and that for an integer k represented as a sum of distinct powers of 2, if $k = 2^{i_0} + 2^{i_1} + \ldots + 2^{i_j}$, then $\Sigma_k^n = \Sigma_{2^{i_0}}^n \cdot \ldots \cdot \Sigma_{2^{i_j}}^n$.

Lemma 2. *The multiplicative complexity of G_n is $n - 1$.*

Proof. Let $n = u_k, u_{k-1}, ..., u_1, u_0$ be the binary representation of n. To compute $G_n(x)$, one first computes the Hamming weight of x, giving $\{\Sigma_{2^k}^n(x) \mid 0 \leq k \leq \lceil \log_2(n+1) \rceil - 1\}$.

This uses $n - H^{\mathbb{N}}(n)$ AND gates, and gives us Σ_2^n directly. Σ_n^n is the product of $\{\Sigma_{2^i}^n \mid u_i = 1\}$, which requires exactly $H^{\mathbb{N}}(n) - 1$ AND gates to compute. Thus, exactly $n - 1$ AND gates are used. The value of G_n is computed with one additional XOR to add Σ_2^n and Σ_n^n. The multiplicative complexity cannot be lower than this since the degree of G_n is n. □

Proposition 1. *The multiplicative complexity of F_n is at least $n - 1$, since the degree is n.*

Lemma 3. *The multiplicative complexity of F_n is $n - 1$ for $3 \leq n \leq 6$.*

Proof. For $n = 3$, $F_n = E_3^3$, which has multiplicative complexity 2. For $n = 4$, $F_n = T_3^4$, which has multiplicative complexity 3. Proofs of the multiplicative complexities of these functions are in [2].

For $n = 5$, compute the Hamming weight of x, giving

$$\{\Sigma_1^5(x), \Sigma_2^5(x), \Sigma_4^5(x)\}.$$

This uses $5 - 2 = 3$ AND gates.

$$\begin{aligned} F_5 &= \Sigma_3^5 \oplus \Sigma_4^5 \oplus \Sigma_5^5 \\ &= (\Sigma_4^5 \oplus \Sigma_2^5) \wedge (\Sigma_4^5 \oplus \Sigma_1^5) \end{aligned}$$

This can be computed using only one additional AND gate.

For $n = 6$, compute the Hamming weight of x, giving

$$\{\Sigma_1^6(x), \Sigma_2^6(x), \Sigma_4^6(x)\}.$$

This uses $6 - 2 = 4$ AND gates.

$$\begin{aligned} F_6 &= \Sigma_3^6 \oplus \Sigma_4^6 \oplus \Sigma_5^6 \oplus \Sigma_6^6 \\ &= (\Sigma_4^6 \oplus \Sigma_2^6) \wedge (\Sigma_4^6 \oplus \Sigma_1^6) \end{aligned}$$

This can be computed using only one additional AND gate. □

Lemma 4. *The multiplicative complexity of F_n is at most $n - H^{\mathbb{N}}(n) + k - 1$, for $k = \lceil \log(n+1) \rceil$.*

Proof. First compute the Hamming weight of the input, that is the functions $\Sigma_{2^i}^n$ for $i = 0, 1, \ldots, k-1$. The function

$$(1 \oplus \Sigma_1^n) \cdot (1 \oplus \Sigma_2^n) \cdot (1 \oplus \Sigma_4^n) \cdot \ldots \cdot (1 \oplus \Sigma_{2^{k-1}}^n)$$

can be computed with $k - 1$ AND gates. This function is equal to

$$1 \oplus \bigoplus_{i=1}^{n} \Sigma_i^n = (1 \oplus x_1)(1 \oplus x_2) \cdot \ldots \cdot (1 \oplus x_n),$$

Since they are both 1 if and only if all input bits are 0. That is F_n can now be obtained without further multiplications since

$$(1 \oplus \Sigma_1^n) \cdot (1 \oplus \Sigma_2^n) \ldots (1 \oplus \Sigma_{2^{k-1}}^n) \oplus 1 \oplus \Sigma_1^n \oplus \Sigma_2^n = F_n$$

□

It turns out that these eight functions have very low annihilator immunity. We consider the variants of F_n functions first and then the variants of G_n.

Lemma 5. *The function* $f = a \oplus b\Sigma_1^n \oplus \bigoplus_{i=3}^n \Sigma_i^n$ *has annihilator immunity at most* 2.

Proof. Let $\tilde{f} = b\Sigma_1^n \oplus \bigoplus_{i=3}^n \Sigma_i^n$, and let $h = 1 \oplus (1 \oplus b)\Sigma_1^n \oplus \Sigma_2^n$ be the algebraic complement of $\tilde{f}$, [29]. Notice that

$$\tilde{f} \oplus h = \bigoplus_{i=1}^{n} \Sigma_i^n \oplus 1 = (1 \oplus x_1)(1 \oplus x_2) \ldots (1 \oplus x_n)$$

which is 1 if and only if $\mathbf{x} = \mathbf{0}$. That is for $\mathbf{x} \neq \mathbf{0}$, $\tilde{f} = h$, so $1 \oplus h$ clearly annihilates $\tilde{f}$ on all non-zero inputs. Since $\tilde{f}(\mathbf{0}) = 0$, h is an annihilator of $\tilde{f}$ with degree 2, so depending on a, h is an annihilator of f. □

Lemma 6. *The function* $f = a \oplus b\Sigma_1^n \oplus \Sigma_2^n \oplus \Sigma_n^n$ *has annihilator immunity at most* 2.

Proof. Let $\mathbf{1}$ denote the all 1 input vector. For some fixed choice of a, and b, depending on n, either $(a \oplus b\Sigma_1^n \oplus \Sigma_2^n)(\mathbf{1}) = 1$ or $(a \oplus b\Sigma_1^n \oplus \Sigma_2^n)(\mathbf{1}) = 0$. In the first case, the function $h = 1 \oplus a \oplus b\Sigma_1^n \oplus \Sigma_2^n$ is an annihilator of f, and otherwise $h = a \oplus b\Sigma_1^n \oplus \Sigma_2^n$ is an annihilator of $f \oplus 1$. And again, clearly there is no annihilator of degree less than 2. □

8 Conclusion

Four nonlinearity concepts are considered and compared, and new relations between them are presented. The four concepts are shown to be distinct; none is subsumed by any of the others.

We are currently extending the ideas present here for cryptanalyzing functions with low multiplicative complexity. It will be interesting to see if using the topology of the circuit for the cryptographic function will lead to useful heuristics for cryptanalytic attacks, especially for variants of hash function with few rounds.

Acknowledgements

We are grateful to Meltem Turan for pointing out some typos in an earlier version of this paper.

References

1. Boyar, J., Damgaard, I., Peralta, R.: Short non-interactive cryptographic proofs. Journal of Cryptology 13, 449–472 (2000)
2. Boyar, J., Peralta, R.: Tight bounds for the multiplicative complexity of symmetric functions. Theor. Comput. Sci. 396(1-3), 223–246 (2008)

3. Boyar, J., Peralta, R., Pochuev, D.: On the multiplicative complexity of Boolean functions over the basis $(\wedge, \oplus, 1)$. Theor. Comput. Sci. 235(1), 43–57 (2000)
4. Canteaut, A., Videau, M.: Symmetric Boolean functions. IEEE Transactions on Information Theory 51(8), 2791–2811 (2005)
5. Carlet, C.: On the degree, nonlinearity, algebraic thickness, and nonnormality of Boolean functions, with developments on symmetric functions. IEEE Transactions on Information Theory 50(9), 2178–2185 (2004)
6. Carlet, C.: Boolean functions for cryptography and error correcting codes. In: Crama, Y., Hammer, P.L. (eds.) Boolean Models and Methods in Mathematics, Computer Science, and Engineering, chap. 8, pp. 257–397. Cambridge, UK: Cambridge Univ. Press (2010)
7. Carlet, C., Dalai, D.K., Gupta, K.C., Maitra, S.: Algebraic immunity for cryptographically significant Boolean functions: Analysis and construction. IEEE Transactions on Information Theory 52(7), 3105–3121 (2006)
8. Courtois, N., Meier, W.: Algebraic attacks on stream ciphers with linear feedback. In: Biham, E. (ed.) EUROCRYPT. LNCS, vol. 2656, pp. 345–359. Springer, Heidelberg (2003)
9. Courtois, N., Hulme, D., Mourouzis, T.: Solving circuit optimisation problems in cryptography and cryptanalysis, e-print can be found at http://eprint.iacr.org/2011/475.pdf
10. Dalai, D.K., Gupta, K.C., Maitra, S.: Results on algebraic immunity for cryptographically significant Boolean functions. In: Canteaut, A., Viswanathan, K. (eds.) INDOCRYPT. LNCS, vol. 3348, pp. 92–106. Springer, Heidelberg (2004)
11. Dalai, D.K., Maitra, S., Sarkar, S.: Basic theory in construction of Boolean functions with maximum possible annihilator immunity. Des. Codes Cryptography 40(1), 41–58 (2006)
12. Demenkov, E., Kulikov, A.S.: An elementary proof of a 3n - o(n) lower bound on the circuit complexity of affine dispersers. In: Murlak, F., Sankowski, P. (eds.) MFCS. LNCS, vol. 6907, pp. 256–265. Springer, Heidelberg (2011)
13. Didier, F.: A new upper bound on the block error probability after decoding over the erasure channel. IEEE Transactions on Information Theory 52(10), 4496–4503 (2006)
14. Dobbertin, H.: Construction of bent functions and balanced Boolean functions with high nonlinearity. In: Preneel, B. (ed.) FSE. LNCS, vol. 1008, pp. 61–74. Springer, Heidelberg (1994)
15. Goldreich, O., Micali, S., Wigderson, A.: How to play any mental game. In: Proceedings of the nineteenth annual ACM symposium on Theory of computing. pp. 218–229. STOC '87, ACM, New York, NY, USA (1987), http://doi.acm.org/10.1145/28395.28420
16. Jukna, S.: Boolean Function Complexity: Advances and Frontiers. Springer Berlin Heidelberg (2012)
17. Kavut, S., Maitra, S., Yücel, M.D.: There exist Boolean functions on n (odd) variables having nonlinearity > $2^{\mathrm{n-1}}$ - $2^{(\mathrm{n-1})/2}$ if and only if n>7. IACR Cryptology ePrint Archive 2006, 181 (2006)
18. Kolesnikov, V., Schneider, T.: Improved garbled circuit: Free XOR gates and applications. In: Aceto, L., Damgård, I., Goldberg, L.A., Halldórsson, M.M., Ingólfsdóttir, A., Walukiewicz, I. (eds.) ICALP (2). LNCS, vol. 5126, pp. 486–498. Springer, Heidelberg (2008)
19. Lupanov, O.: On rectifier and switching-and-rectifier schemes. Dokl. Akad. 30 Nauk SSSR 111, 1171-1174. (1965)

20. Maitra, S., Sarkar, P.: Maximum nonlinearity of symmetric Boolean functions on odd number of variables. IEEE Transactions on Information Theory 48(9), 2626–2630 (2002)
21. McFarland, R.L.: Sub-difference sets of Hadamard difference sets. J. Comb. Theory, Ser. A 54(1), 112–122 (1990)
22. Nechiporuk, E.I.: On the complexity of schemes in some bases containing nontrivial elements with zero weights (in Russian). Problemy Kibernetiki 8, 123–160 (1962)
23. Nielsen, J.B., Nordholt, P.S., Orlandi, C., Burra, S.S.: A new approach to practical active-secure two-party computation. In: Safavi-Naini, R., Canetti, R. (eds.) CRYPTO. LNCS, vol. 7417, pp. 681–700. Springer, Heidelberg (2012)
24. O'Donnell, R.: Analysis of Boolean Functions. Book draft. Available at www.analysisofbooleanfunctions.org (2012)
25. Rodier, F.: Asymptotic nonlinearity of Boolean functions. Des. Codes Cryptography 40(1), 59–70 (2006)
26. Rothaus, O.S.: On "bent" functions. J. Comb. Theory, Ser. A 20(3), 300–305 (1976)
27. Savický, P.: On the bent Boolean functions that are symmetric. Eur. J. Comb. 15(4), 407–410 (1994)
28. Schnorr, C.P.: The multiplicative complexity of Boolean functions. In: Mora, T. (ed.) AAECC. LNCS, vol. 357, pp. 45–58. Springer, Heidelberg (1988)
29. Zhang, X., Pieprzyk, J., Zheng, Y.: On algebraic immunity and annihilators. Information Security and Cryptology–ICISC 2006 pp. 65–80 (2006)
30. Zhegalkin, I.I.: On the technique of calculating propositions in symbolic logic. Matematicheskii Sbornik 43, 9–28 (1927)