Algebraic nonlinearity and its applications to cryptography

Luke O'Connor
Department of Computer Science
University of Waterloo, Ontario, Canada, N2L 3G1
email: ljpoconn@watmath.uwaterloo.ca

Andrew Klapper*
Department of Computer Science
University of Kentucky, Kexington, Kentucky, USA, 40506-0027
and Computer Science Department
University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2
email: klapper@cs.uky.edu

Abstract

The algebraic nonlinearity of an n-bit boolean function is defined as the degree of the polynomial $f(X) \in Z_2[x_1, x_2, \ldots, x_n]$ that represents f. We prove that the average degree of an ANF polynomial for an n-bit function is $n + o(1)$. Further for a balanced n-bit function, any subfunction obtained by holding less than $n - \lceil \log n \rceil - 1$ bits constant is also expected to be nonaffine. A function is partially linear if $f(X)$ has some indeterminates that only occur in terms bounded by degree 1. Boolean functions which can be mapped to partially linear functions via a linear transformation are said to have a linear structures, and are a potentially weak class of functions for cryptography. We prove that the number of n-bit functions that have a linear structure is asymptotic $(2^n - 1) \cdot 2^{2^{n-1}+1}$.

Keywords: linearity, partial linearity, linear structures.

*Project sponsored in part by NSERC operating grant OGP0121648, and the National Security Agency under Grant Number MDA904-91-H-0012. The United States Government is authorized to reproduce and distribute reprints notwithstanding any copyright notation hereon.

1 Introduction

Encryption mappings, particularly product ciphers, are often designed to satisfy a set of chosen criteria which have been established either formally or empirically as essential to the security of the cipher [25]. Two basic criteria are due to Shannon [29] who suggested that a product cipher should be constructed using the notions of *diffusion* and *confusion*. Diffusion refers to the dissipation of the statistical properties of the plaintext, while confusion refers to the internal operations of the cipher that produce complex relations between the plaintext, key and ciphertext.

More recently these notions have been refined by modeling product ciphers, and components thereof, using boolean functions. For example, if a ciphertext bit c_i is described by the boolean function f_i then it is generally accepted [1, 6, 14, 17, 20, 21] that each f_i should possess a combination of the following properties: balance, nonlinearity [14], nondegeneracy/completeness [10, 11], correlation immunity [28], satisfy the strict avalanche criterion [31], or be bent [23]. These properties may be collectively referred to as *nonlinearity criteria* [14, 25] and can be extended in several natural ways. Consider defining a class P of boolean functions which are known to have a cryptographic weakness, and then selecting functions which are optimized to be maximally 'dissimilar' from every function in P. One measure of dissimilarity is to interpret an n-bit function as a vector with 2^n coordinates, and use the Hamming distance metric. Then for example, functions may be chosen so that they achieve a maximum distance from all functions that are affine [14] or have linear structures [4] (defined below). Also we may consider a nonlinearity criterion to be robust if it is invariant under certain simple mappings such as affine transformations. Meier and Staffelbach [14] have shown that the distance to the set of linear functions, and the nonlinear order of a function [26] are both invariant under nonsingular linear transformation.

A property P, such as nonlinearity, in a function may be considered stronger in the function if P is still retained when certain subsets of the input bits are held constant. This has been referred to as the *higher order characteristics* of property P, and for example, has been considered in extending the notions of correlation immunity [28] and the strict avalanche criterion [7]. Here we may assume that an assignment to a subset of the input bits represents any partial knowledge that a cryptanalyst may have about the key or plaintext, and if the property P is preserved under this assignment, the characteristics of the function are not biased by this partial knowledge.

For product ciphers, nonlinearity criteria are typically applied to the construction of the S-boxes to be used in the round function [1, 6, 17, 20]. The nonlinearity of a product cipher depends directly on the selection of these S-boxes since, typically, the S-boxes are the only nonaffine component of the cipher; in particular, if the S-boxes are affine then the entire mapping is then affine (as is the case for DES). In 1985 Reeds and Manferdelli [22] devised an attack they called *cryptosystem factorization*. The idea is that there may exist separate affine functions for the plaintext, ciphertext and key such that in the mapped domains the dimensionality of the keyspace has been reduced (that is, in the mapped domain certain key bits are degenerate). If this was the case then the cost of exhaustive search of the keyspace would be reduced. For DES, Reeds and Manferdelli [22] showed that no such factorization of the round mapping exists.

Generalizing these ideas, Chaum and Evertse [4] defined *linear structures* and devised an attack on DES which is less costly than exhaustive search when DES is restricted to fewer than 8 rounds. An n-bit function $f : Z_2^n \to Z_2$ is said to have a linear structure $b \neq \mathbf{0} \in Z_2^n$ if and only if $f(X) \oplus f(X+b)$ is independent of X. Subsequently Lai [13] has shown that if f has k linearly independent vectors $b_1, b_2, \ldots, b_k$ that are linear structures, then f can be mapped to g via a linear transformation where

$$g = x_1 m_1 + x_2 m_2 + \cdots + x_k m_k + g'(x_{k+1}, x_{k+2}, \ldots, x_n). \tag{1}$$

The cryptanalyst may be able to take advantage of the linear structures in f if some of the m_i in equation (1) are zero, thus eliminating the influence of some variables (possibly key bits) on the ciphertext. Meier and Staffelbach [14] have shown that for even n the bent functions attain the maximum distance from the class of n-bit functions that have linear structures.

Differential cryptanalysis [3] can be seen as an extension of the ideas of attacks based on the presence of linear structures [18]. We may alternately state that $b \neq \mathbf{0} \in Z_2^n$ is a linear structure of an n-bit function f if and only if

$$\sum_{X \in Z_2^n} \hat{f}(X) \cdot \hat{f}(X+b) = \pm 2^n$$

where $\hat{f}(X) = (-1)^{f(X)}$. Thus inputs of difference b result in an output of difference zero or one with probability 1. In differential cryptanalysis, it is only required that inputs of difference ΔX lead to a known difference ΔY with high probability, or with a probability that noticeably exceeds the mean. Evertse [5] defined a function f as having a *50%-linear structure* with respect to $b \neq \mathbf{0}$ if

$$\sum_{X \in Z_2^n} \hat{f}(X) \cdot \hat{f}(X+b) = 0.$$

Evertse was sceptical that S-boxes could be designed which satisfied this property for each $b \in Z_2^n$, and for each output bit of an S-box. Meier and Staffelbach [14] later defined a function f as being *perfect nonlinear* if for all $b \neq \mathbf{0} \in Z_2^n$, b is a 50%-linear structure for f. Equivalently, we say that f has maximum distance to the class of linear structures. If an n-bit function f is perfect nonlinear, then given $\hat{f}(X) \cdot \hat{f}(X+b)$, all $b \neq \mathbf{0} \in Z_2^n$ are equally likely to have produced the output difference. This suggests that perfect nonlinear linear functions are a useful class of functions for constructing mappings that are resistant to differential attacks.

1.1 Results

As most functions are nonlinear we may then consider methods of *ranking* functions according to their nonlinearity, which will requires a *measure of nonlinearity*. There are two accepted measures of nonlinearity, interpreted as an *algebraic* measure and a *functional* measure. It is well-known that any n-bit boolean function f can be expressed as a polynomial in $Z_2[x_1, x_2, \ldots, x_n]$ whose degree in each variable is at most one. This polynomial is known as the algebraic normal form (ANF) of f. We shall abuse notation slightly and use the name of a function to denote its ANF. The algebraic measure of nonlinearity of a function is simply the degree of its ANF; on the other hand, the functional measure of nonlinearity is the minimal distance from the function to the set of all affine functions [14, 20].

We begin by proving (Theorem 2.1) that the average degree of an ANF polynomial is $n + o(1)$, which implies that a randomly selected function will have high algebraic nonlinearity. Even though a function may have a low probability of being affine it may be possible to induce affinity by holding constant a subset of the input bits. We will prove (Theorem 2.3) that for a balanced function (typically, the boolean functions that are used cryptographically are balanced), at least $n - \lceil \log n \rceil - 1$ bits must be assigned before any affine subfunction is expected to exist. Thus a function *tends to* remain nonaffine even when a large number of input variables are assigned. This result has implications for ciphertext-only attacks. For any fixed key K the boolean equations which describe the ciphertext in terms of the plaintext are balanced, since the encryption function is invertible. Then if the block length is n, on average $n - \lceil \log n \rceil - 1$ bits must be determined before the remaining unknown plaintext bits exhibit degenerate relations.

A natural extension of algebraic nonlinearity is to consider *partial linearity*. A function f will be said to be partially linear if there exists an indeterminate x_i which only occurs as a linear term in the ANF of f. From equation (1) we see that if f has k linearly independent structures then f can be mapped linearly to a function g that is partially linear in k variables. Linear structures encompass a very broad notion of algebraic linearity, since a function f with no linear structures cannot be mapped to a partially linear function via a linear transformation. In §3 we use the Möbius inversion formula, applied to the lattice of vector subspaces of a vector space, to enumerate the set of n-bit functions that possess linear structures. We prove (Theorem 3.1) that the number of functions with a linear structure approaches $(2^n - 1) \cdot 2^{2^{n-1}+1}$. For example, the probability that a randomly selected 6-bit function has a linear structure is less than 10^{-7}.

2 Properties of ANF polynomials

There are several normal forms for boolean functions [10, 27] and of special interest to cryptography is the Algebraic Normal Form (ANF) [26, p.130], also known as the Ring Sum Expansion (RSE) [27, p.19]. The set of all ANFs of functions, denoted by $Z_2^A[x_1, x_2, \ldots, x_n]$, is given as

$$Z_2^A[x_1, x_2, \ldots, x_n] \;=\; \left\{ \sum_{S \subseteq [n]} a_S \prod_{j \in S} x_j \right\},$$

where $a_S \in Z_2$, $[n] = \{1, 2, \ldots, n\}$. The set of n-bit affine functions $\mathcal{A}^n$ is exactly those functions for which the total degree of f is bounded by 1. Let $\deg(f, n)$ be the degree of f. The main result of this section is that the expected degree of f is larger than 1, even when a significant number of indeterminates are assigned values.

Theorem 2.1 Assuming the uniform distribution on $Z_2^A[x_1, x_2, \ldots, x_n]$

$$\begin{aligned} \mathbf{E}[\deg(f,n)] &= n - \frac{1}{2} + \Theta\left(\frac{1}{2^n}\right) \\ \mathbf{Var}[\deg(f,n)] &= \frac{1}{4} + \Theta\left(\frac{1}{2^n}\right). \end{aligned}$$

Proof. Let $B(m,k) = \sum_{0 \le j \le k} \binom{m}{j}$ be the sum of the first $k+1$ binomial coefficients, $0 \le k \le m$. Observe that

$$2^{B(n,n-2)} = 2^{2^n - n - 1},$$

from which it follows that

$$\begin{aligned} 2^{2^n-n-1} &\le \sum_{i=0}^{n-2} 2^{B(n,i)} \\ &\le 2^{2^n-n-1} + \sum_{i=0}^{n-3} 2^{B(n,n-3)} \\ &= 2^{2^n-n-1} + (n-3) \cdot 2^{2^n - \frac{n(n+1)}{2} - 1} \\ &= 2^{2^n-n-1} \cdot \left(1 + (n-3) \cdot 2^{-\frac{n(n-1)}{2}}\right) \\ &\le 2^{2^n-n} \end{aligned}$$

for n sufficiently large. Therefore

$$\mathbf{E}[\deg(f,n)] \;=\; \sum_{0 \le i \le n} i \cdot \Pr(f \text{ has degree } i\,)$$

$$
\begin{aligned}
&= \sum_{0 \le i \le n} i \cdot \frac{2^{B(n,i-1)} \cdot (2^{\binom{n}{i}} - 1)}{2^{2^n}} \\
&= 2^{-2^n} \cdot \left[\sum_{i=1}^{n} i \cdot 2^{B(n,i)} - i \cdot 2^{B(n,i-1)} \right] \\
&= 2^{-2^n} \cdot \left[n \cdot 2^{2^n} - 2^{2^n - 1} - \sum_{i=0}^{n-2} 2^{B(n,i)} \right] \\
&= n - \frac{1}{2} + \Theta\Big(\frac{1}{2^n}\Big).
\end{aligned}
$$

Similarly, the variance $\mathbf{Var}[\deg(f,n)]$ may be computed as

$$
\begin{aligned}
\mathbf{Var}[\deg(f,n)] &= 2^{-2^n} \cdot \left[\sum_{i=1}^{n} i^2 \cdot 2^{B(n,i)} - i^2 \cdot 2^{B(n,i-1)} \right] - \mathbf{E}[\deg(f,n)]^2 \\
&= 2^{-2^n} \cdot \left[n^2 \cdot 2^{2^n} - \sum_{i=0}^{n-1} (2i+1) \cdot 2^{B(n,i)} \right] - \mathbf{E}[\deg(f,n)]^2 \\
&= n^2 - \frac{2n-1}{2} - 2^{-2^n} \cdot \sum_{i=0}^{n-2} (2i+1) \cdot 2^{B(n,i)} - \left(n - \frac{1}{2} - 2^{-2^n} \sum_{i=0}^{n-2} 2^{B(n,i)} \right)^2 \\
&= \frac{1}{4} - 2^{-2^n} \cdot \sum_{i=0}^{n-2} (2i+1) \cdot 2^{B(n,i)} + (2n-1) \cdot 2^{-2^n} \sum_{i=0}^{n-2} 2^{B(n,i)} \\
&\quad - 2^{-2^{n+1}} \cdot \left(\sum_{i=0}^{n-2} 2^{B(n,i)} \right)^2 \\
&= \frac{1}{4} + 2^{1-2^n} \cdot \sum_{i=0}^{n-2} (n-i-1) \cdot 2^{B(n,i)} - 2^{-2^{n+1}} \cdot \left(\sum_{i=0}^{n-2} 2^{B(n,i)} \right)^2 . \qquad (2)
\end{aligned}
$$

To estimate the second expression in equation (2), we have

$$
\begin{aligned}
2^{2^n - n - 1} &\le \sum_{i=0}^{n-2} (n-i-1) \cdot 2^{B(n,i)} \\
&= 2^{B(n,n-2)} + \sum_{i=0}^{n-3} (n-i-1) \cdot 2^{B(n,i)} \\
&\le 2^{2^n - n - 1} + (n-2)^2 \cdot 2^{B(n,n-3)} \\
&= 2^{2^n - n - 1} + (n-2)^2 \cdot 2^{2^n - \frac{n(n+1)}{2} - 1} \\
&= 2^{2^n - n - 1} \cdot \Big(1 + (n-2)^2 \cdot 2^{\frac{n(n-1)}{2}} \Big) \\
&\le 2^{2^n - n}.
\end{aligned}
$$

To estimate the last expression in equation (2), we have

$$2^{-2^{n+1}} \cdot \left(\sum_{i=0}^{n-2} 2^{B(n,i)} \right)^2 = \Theta\left(\frac{2^{2^n - n}}{2^{2^{n+1}}} \right) = \Theta\left(\frac{1}{2^{2^n+n}} \right).$$

It follows that

$$\mathbf{Var}[\deg(f,n)] \;=\; \frac{1}{4} + \Theta\left(\frac{1}{2^n}\right).$$

□

It can also be shown using the binomial theorem that the expected number of terms in (the ANF of) f is 2^{n-1} with a variance of 2^{n-2}. Even though a function may have a low probability of being affine, it may be possible to induce affiness by holding constant a subset of the plaintext bits. An order d subfunction f', $0 \le d \le n$, of an n-bit function f, is any $(n-d)$-bit function obtained by holding d inputs of f constant. Let $g \in Z_2^A[y_1, y_2, \ldots, y_{n-d}]$ be an $(n-d)$-bit function where $0 \le d \le n$ and $\{y_1, y_2, \ldots, y_{n-d}\} \subseteq \{x_1, x_2, \ldots, x_n\}$. We say that f has g as a subfunction if there exists an order d subfunction f' of f such that $f' = g$. We show that for a balanced function f, on average $n - \lceil \log n \rceil - 1$ bits must be assigned (known) before any affine subfunction is induced, assuming all balanced functions f are equally likely.

The proof of the next theorem is similar to the work of Mileto and Putzolu [15, 16] on determining the average number of prime implicants in a boolean function. Observe that a prime implicant of a boolean function f corresponds to a constant subfunction of f. Let the weight of an n-bit function f be defined as $|\{X : f(X) = 1,\ X \in Z_2^n\}|$. Thus f is balanced if its weight is 2^{n-1}. Also define $\binom{m}{k} = 0$ if $k < 0$.

Theorem 2.2 Let f be an n-bit function of weight k, $0 \le k \le 2^n$. Let f have $H_k^n(f,r)$ affine subfunctions of order $n-r$, $0 \le r < n$. Assuming all weight k functions are equally likely, we have

$$\mathbf{E}[H_k^n(f,r)] \;=\; \frac{\binom{n}{r} \cdot 2^{n-r}}{\binom{2^n}{k}} \cdot \left[(2^{r+1} - 2) \cdot \binom{2^n - 2^r}{k - 2^{r-1}} + \binom{2^n - 2^r}{k} + \binom{2^n - 2^r}{k - 2^r} \right]. \quad (3)$$

Proof. For $F(n,k) = \binom{2^n}{k}$, let $f_1, f_2, \ldots, f_{F(n,k)}$ be the n-bit functions of weight k. Also, for a given n-bit function f let $f_1', f_2', \ldots, f_{C(n,r)}'$ be the $C(n,r)$ order $(n-r)$ subfunctions of f where $C(n,r) = \binom{n}{r} \cdot 2^{n-r}$. Let $\mathcal{A}^r = \{\, g_i^r \mid 1 \le i \le 2^{r+1} \,\}$ be an arbitrary explicit enumeration. By definition we have that

$$H_k^n(f,r) \;=\; \sum_{c=1}^{C(n,r)} \sum_{i=1}^{2^{r+1}} [g_i^r = f_c'],$$

where $[\cdot]$ is a boolean predicate evaluating to 0 or 1. It then follows that

$$
\begin{aligned}
\mathbf{E}[H_k^n(f,r)] &= \frac{1}{F(n,k)} \cdot \sum_{j=1}^{F(n,k)} H_k^n(f_j,r) \\
&= \frac{1}{F(n,k)} \cdot \sum_{j=1}^{F(n,k)} \sum_{c=1}^{C(n,r)} \sum_{i=1}^{2^{r+1}} [g_i^r = f'_{j,c}] \\
&= \frac{1}{F(n,k)} \cdot \sum_{i=1}^{2^{r+1}} \sum_{c=1}^{C(n,r)} \sum_{j=1}^{F(n,k)} [g_i^r = f'_{j,c}]. \qquad (4)
\end{aligned}
$$

The inner summations (over c and j) of equation (4) give the number of functions that have the affine function g_i^r as a fixed order $(n-r)$ subfunction. We want to find the average number of times a function of weight k contains a fixed g_i^r of weight w as a subfunction. This can be computed as the total number of instances of g_i^r occuring as a subfunction of any weight k function, divided by the number $F(n,k)$ of weight k functions.

To realize g_i^r as a subfunction of a function f, we first choose $\{y_1,\cdots,y_r\} \subseteq \{x_1,\cdots,x_n\}$ for the r bits of g_i^r, and values for the remaining $n-r$ bits of f. There are $\binom{n}{r} \cdot 2^{n-r} = C(n,r)$ ways of making these choices. We require that f induce g_i^r when these values are assigned, and this determines the value of f on 2^r points. To determine f on the remaining $2^n - 2^r$ points, and guarantee that f has weight k, we must pick the $k-w$ points where f will take the value one. There are

$$\binom{2^n - 2^r}{k-w}$$

such choices. This gives

$$C(n,r) \cdot \binom{2^n - 2^r}{k-w}$$

ways of realizing g_i^r as a subfunction of a weight k function. Thus the average number of times a function of weight k contains g_i^r as a subfunction is

$$\frac{C(n,r)}{F(n,k)} \cdot \binom{2^n - 2^r}{k-w}.$$

It follows that

$$\mathbf{E}[H_k^n(f,r)] = \frac{C(n,r)}{F(n,k)} \cdot \left[(2^{r+1}-2)\binom{2^n-2^r}{k-2^{r-1}} + \binom{2^n-2^r}{k} + \binom{2^n-2^r}{k-2^r} \right]$$

where the 3 binomial coefficients correspond to the $(2^{r+1}-2)$ balanced affine functions, and the 2 constant affine functions, respectively. $\square$

Consider the set Γ^n of all n-bit functions, with the uniform distribution on this set. Let $L_n(f) : \Gamma^n \to \{0, 1, \ldots, n-1\}$ be a random variable, such that if $L_n(f) = d$ then to induce an affine subfunction in f it is necessary and sufficient to assign d variables. That is, if a function f depends on the variables $V = \{ x_1, x_2, \ldots, x_n \}$, then $\Pr(L_n(f) = d)$ is the probability that there exists a set $V' \subseteq V$, $|V'| = d$, such that it is possible to induce affiness in a subfunction of f by making an assignment to the variables of V', and there is no smaller set with this property. Observe that $L_n(f) \leq n-1$ since all order $(n-1)$ subfunctions must be affine.

It follows that $\mathbf{E}[L_n(f)] = \sum_{d=0}^{n-1} d \cdot \Pr(L_n(f) = d)$ gives the expected number of variables that must be assigned before an affine subfunction is induced, where the expectation is taken over all possible n-bit functions.

Theorem 2.3 For large n, assuming all balanced functions are equally likely, $\mathbf{E}[L_n(f)] \geq n - \lceil \log n \rceil - 1 + o(1)$. Moreover, $\mathbf{Var}[L_n(f)] \leq (\lceil \log n \rceil + 1)^2/4 + o(1)$.

Proof. To simplify notation, we let $r = n - d$. Observe that for balanced functions $\Pr[L_n(f) = n - r] \leq \mathbf{E}[H^n_{2^{n-1}}(f, r)]$. We begin by obtaining an asymptotic estimate of $\mathbf{E}[H^n_{2^{n-1}}(f, r)]$. From equation (3) we have that

$$\Pr(L_n(f) = n - r) \;\leq\; \mathbf{E}[H^n_{2^{n-1}}(f, r)] \;<\; \frac{\binom{n}{r} \cdot 2^{n+1} \cdot \binom{2^n - 2^r}{2^{n-1} - 2^{r-1}}}{\binom{2^n}{2^{n-1}}}. \tag{5}$$

We prove that $\mathbf{E}[H^n_{2^{n-1}}(f, r)] << 1$ for $r > \lceil \log n \rceil + 1$. Using a sharp form of Stirling's formula, such as that found in Knuth, vol. 1 [12, p. 111], one sees that

$$\sqrt{2\pi m} \left(\frac{m}{e}\right)^m \leq m! \leq 2\sqrt{\pi m} \left(\frac{m}{e}\right)^m,$$

for any m. It follows that we can estimate the central binomial coefficient by

$$\left(\frac{1}{\pi m}\right)^{1/2} 2^{2m-1} < \binom{2m}{m} < \left(\frac{2}{\pi m}\right)^{1/2} 2^{2m}.$$

Therefore

$$\begin{aligned}
\mathbf{E}[H^n_{2^{n-1}}(f, r)] \;&<\; \frac{\binom{n}{r} \cdot 2^{n+1} \cdot \left(\frac{1}{\pi \cdot (2^n - 2^r)}\right)^{\frac{1}{2}} \cdot 2^{2^n - 2^r + 1}}{\left(\frac{2}{\pi \cdot 2^n}\right)^{\frac{1}{2}} \cdot 2^{2^n - 1}} \\
&= \frac{\binom{n}{r} \cdot 2^{n + \frac{n}{2} + \frac{5}{2}}}{2^{2^r} \cdot (2^n - 2^r)^{\frac{1}{2}}}.
\end{aligned}$$

When $r \geq \lceil \log n \rceil + 2$ we have

$$\begin{aligned}
\mathbf{E}[H^n_{2^{n-1}}(f,r)] &< \frac{\binom{n}{r} \cdot 2^{n+\frac{n}{2}+\frac{5}{2}}}{2^{2^{\lceil \log n \rceil + 2}} \cdot (2^n - 4n)^{\frac{1}{2}}} \\
&< \frac{\binom{n}{r} \cdot 2^{\frac{3n}{2}}}{2^{4n}} \\
&< \binom{n}{r} \cdot 2^{-2n}
\end{aligned}$$

for n sufficiently large. Then observe that

$$\begin{aligned}
\sum_{r=\lceil \log n \rceil + 2}^{n} r \cdot \Pr(L_n(f) = n-r) &\leq \sum_{r=\lceil \log n \rceil + 2}^{n} r \cdot \mathbf{E}[H^n_{2^{n-1}}(f,r)] \\
&< \sum_{r=\lceil \log n \rceil + 2}^{n} r \cdot \binom{n}{r} \cdot 2^{-2n} \\
&\leq \sum_{r=0}^{n} n \cdot \binom{n}{r} \cdot 2^{-2n} \\
&= n \cdot 2^n \cdot 2^{-2n} \\
&= o(1).
\end{aligned}$$

It follows that for large n (recalling that $d = n - r$)

$$\begin{aligned}
\mathbf{E}[L_n(f)] &= \sum_{d=0}^{n-1} d \cdot \Pr(L_n(f) = d) \\
&= n - \sum_{r=1}^{n} r \cdot \Pr(L_n(f) = n-r) \\
&= n + o(1) - \sum_{r=1}^{\lceil \log n \rceil + 1} r \cdot \Pr(L_n(f) = n-r) \\
&\geq n - \lceil \log n \rceil - 1 + o(1).
\end{aligned}$$

Next we consider the variance:

$$\begin{aligned}
\mathbf{Var}[L_n(f)] &= \mathbf{E}[L_n(f)^2] - \mathbf{E}[L_n(f)]^2 \\
&= \sum_{d=0}^{n-1} d^2 \cdot \Pr(L_n(f) = d) - \left(\sum_{d=0}^{n-1} d \cdot \Pr(L_n(f) = d) \right)^2 \\
&= \sum_{r=1}^{n} (n^2 - 2nr + r^2) \cdot \Pr(L_n(f) = n-r) - \left(n - \sum_{r=1}^{n} r \cdot \Pr(L_n(f) = n-r) \right)^2
\end{aligned}$$

$$= \sum_{r=1}^{n} r^2 \cdot \Pr(L_n(f) = n-r) - \left(\sum_{r=1}^{n} r \cdot \Pr(L_n(f) = n-r)\right)^2$$

$$= \sum_{r=1}^{\lceil \log n \rceil + 1} r^2 \cdot \Pr(L_n(f) = n-r) - \left(\sum_{r=1}^{\lceil \log n \rceil + 1} r \cdot \Pr(L_n(f) = n-r)\right)^2 + o(1).$$

An expression of the form $\sum_{r=1}^{k} r^2 x_r - (\sum_{r=1}^{k} r x_r)^2$ is maximized when $x_1 = \cdots = x_{k-1} = 0$ and $x_k = 1/2$, and its value at this point is $k^2/4$. Thus the variance is bounded above by $(\lceil \log n \rceil + 1)^2 + o(1)$.

This completes the proof of the theorem. □

Thus at least $n - \lceil \log n \rceil - 1$ bits must set before an affine subfunction is expected to be induced in a balanced function. Table 1 shows bounds on $\sum_{r \geq \lceil \log n \rceil + 2} (n-r) \cdot \Pr(L_n(f) = r)$ as computed from equation (5). It is clear that the tail of the expectation $\mathbf{E}[L_n(f)]$ beyond $r = \lceil \log n \rceil + 2$ is approaching zero rapidly.

n	$\lceil \log n \rceil + 2$	$\mathbf{E}[H^n_{2^{n-1}}(f, \lceil \log n \rceil + 2))]$	$\sum_{r \geq \lceil \log n \rceil + 2} (n-r) \cdot \mathbf{E}[H^n_{2^{n-1}}(f, r)]$
6	5	0.25189×10^{-6}	0.25189×10^{-6}
7	5	0.14443×10^{-5}	0.28887×10^{-5}
8	5	0.71356×10^{-5}	0.21406×10^{-4}
9	6	0.49845×10^{-14}	0.14953×10^{-13}
10	6	0.24078×10^{-13}	0.96315×10^{-13}

Table 1: Bounds on the tail of $\mathbf{E}[L_f(n)]$ for $n = 6, 7, \ldots, 10$

3 Partial linearity

Functions that exhibit properties common to linear functions are also considered to be cryptographically weak. Observe that the simultaneous complementation of a subset of the input variables causes the value of a linear function to always change (from the original value before complementation) or to never change. If such a subset of the input parameters exists for an arbitrary function, then the function is said to be *partially linear*. The class of functions that possess linear structures are exactly those functions that can be mapped to a partially linear function via a linear transformation (see Theorem 3.1).

We may identify partial linearity as a property of the ANF for a function. Let f be an n-bit function. Then f is said to be *partially linear*, or *p-linear*, if there exists a subset

$Y = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$, $1 \leq k \leq n$, of the variables such that

$$f(X) \;=\; g(x'_1, x'_2, \ldots, x'_{n-k}) \;+ \sum_{1 \leq j \leq k} m_j x_{i_j} \tag{6}$$

where $\{x'_1, x'_2, \ldots, x'_{n-k}\} = \{x_1, x_2, \ldots, x_n\} - Y$, $m_j \in Z_2$, $1 \leq j \leq k$. These functions were previously studied by Beale and Monaghan [2], where they were called *linear-in* functions. For $X = (x_1, x_2, \ldots, x_n) \in Z_2^n$, $X + e_i$ complements the *ith* coordinate of X, where e_i is the *ith* unit vector. Then when an n-bit function f is considered as $f : Z_2^n \to Z_2$, f is p-linear in k variables if and only if there exists a set $B = \{b_1, b_2, \ldots, b_k\} \subseteq \{e_1, e_2, \ldots, e_n\}$ such that for each $b_i \in B$, $f(X) \oplus f(X + b_i)$ is independent of X.

Linear structures are a natural extension of p-linearity, where the set B is an arbitrary subset of Z_2^n. If $f(x_1, x_2, x_3) = \overline{x_1}\,\overline{x_2} x_3 + x_1 \overline{x_2}\,\overline{x_3}$, it can be verified that $f(x_1, x_2, x_3) = f(\overline{x_1}, x_2, \overline{x_3})$ for all values of x_1, x_2, x_3. In other words, $f(X + 101) \oplus f(X)$ is invariant, and $b = 101$ is said to be a *linear structure* of f. It can be shown (see Lemma 3.1) that if f has a linear structure, then there is a linear transformation M that maps f onto a partially linear function.

Let $\mathcal{PL}^n$ be the set of n-bit p-linear functions, and let $\mathcal{LS}^n$ be the set of n-bit functions that have linear structures. It follows from our previous discussion that $\mathcal{A}^n \subset \mathcal{PL}^n \subset \mathcal{LS}^n$. We observe that the set of n-bit degenerate functions [10, 19] is contained in $\mathcal{PL}^n$ as degeneracy is a special case of p-linearity (a subset of the m_i in equation (6) are zero). The cryptanalytic value of linear structures lies in their potential to map a nonlinear function to a degenerate function via a linear transformation, which may reduce the size of the keyspace. Linear structures were introduced by Chaum and Evertse [4] who cryptanalyzed a version of DES restricted to fewer than 8 rounds.

Linear structures encompass a very broad notion of algebraic linearity, and we are interested in determining the probability that a function has a linear structure. If $f \notin \mathcal{LS}^n$ then f cannot be mapped to a p-linear function via a linear transformation, and is considered strongly nonlinear by the algebraic measure of nonlinearity. The set of p-linear functions $\mathcal{PL}^n$ was enumerated by Beale and Monaghan [2] using the inclusion-exclusion principle [9]:

$$|\mathcal{PL}^n| \;=\; \sum_{k=1}^{n} (-1)^{k+1} \cdot \binom{n}{k} \cdot 2^k \cdot 2^{2^{n-k}}.$$

In §3 we will use the Möbius inversion formula applied to the lattice of vector subspaces of a vector space to enumerate the set of n-bit functions that possess linear structures, and prove (Theorem 3.1) that $|\mathcal{LS}^n| \sim (2^n - 1) \cdot 2^{2^{n-1}+1}$.

3.1 The number of functions with linear structures

The relation between p-linearity and linear structures is given in the next lemma.

Lemma 3.1 (Lai [13]) Let $b_1, b_2, \ldots, b_k$ be a set of linearly independent linear structures for the n-bit function f, $1 \leq k \leq n$. There exists a nonsingular $n \times n$ matrix M with coefficients over Z_2 such that if $g(x_1, x_2, \ldots, x_n) = f((x_1, x_2, \ldots, x_n)M)$ then in ANF $g(x_1, x_2, \ldots, x_n)$ is given as

$$g(X) \quad = \quad x_1 m_1 + x_2 m_2 + \cdots + x_k m_k + g'(x_{k+1}, x_{k+2}, \ldots, x_n) \tag{7}$$

where $m_i = f(b_i) \oplus f(\mathbf{0}) \in Z_2$, $1 \leq i \leq k$. □

Corollary 3.1 Let $b_1, b_2, \ldots, b_k$ be a set of linearly independent vectors. There are $2^{2^{n-k}+k}$ n-bit functions for which $b_1, b_2, \ldots, b_k$ are linear structures.

Proof. By Lemma 3.1 let $b_i = e_i$, $1 \leq i \leq k$, without loss of generality. However it follows from equation (7) that there are 2^k ways to choose the m_i, and $2^{2^{n-k}}$ ways to choose the $(n-k)$-bit function g. □

Thus if f is a function that has linear structures $b_1, b_2, \ldots, b_k$, an appropriate basis change for Z_2^n transforms f into a p-linear function.

Example 3.1 The 4-bit function f has $b = 1110$ as a linear structure where

$$f(X) \quad = \quad x_2 + x_1x_2 + x_1x_3 + x_2x_3 + x_3x_4 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4.$$

Define M as the matrix

$$M \quad = \quad \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

If we have that $g(x_1, x_2, \ldots, x_n) = f((x_1, x_2, \ldots, x_n)M)$, then

$$g(X) \quad = \quad x_3 + x_2x_4 + x_3x_4 + x_2x_3x_4.$$

As the first column of M is b, then e_1 is a linear structure in g, and g is degenerate in x_1 as $f(b) = f(\mathbf{0}) = 0$. □

The next lemma is easily proven.

Lemma 3.2 Let f be an n-bit function such that b_1 and b_2 are linear structures for f. Then $b = b_1 + b_2$ is also a linear structure for f. □

Consider counting the number of functions f for which all $b \in B = \{b_1, b_2, \ldots, b_k\}$ are nonzero linear structures. Let the rank d of B be defined as the smallest integer where there exist d vectors $b'_1, b'_2, \ldots, b'_d \in B$, such that $b \in B$ can be written as linear combination of these vectors. Equivalently, d is the dimension of the space spanned by B. Observe that d is bound as $\lceil \log(k+1) \rceil \le d \le k$ since $k - d \le 2^d - d - 1$. From Lemma 3.2, if the b'_i are linear structures for a function f, then every $b \in B$ is a linear structure of f. Thus our problem reduces to counting the number of functions for which $b'_1, b'_2, \ldots, b'_d$ are linear structures. Without loss of generality we may assume that $b'_i = e_i$, $1 \le i \le d$. Corollary 3.1 indicates that the number of linear structures is an exponentially decreasing function of d, the rank. We will use this observation to bound $|\mathcal{LS}^n|$ via the Möbius inversion formula.

Theorem 3.1 $\lim_{n\to\infty} |\mathcal{LS}^n|/((2^n - 1) \cdot 2^{2^{n-1}+1}) = 1$.

Proof. For $b \neq \mathbf{0} \in Z_2^n$, let $P(b)$ be the set functions that have b as a linear structure. Also, for a vector subspace $V \subseteq Z_2^n$, let $P(V) = \cap_{b \in V} P(b)$; that is, $P(V)$ is the set of functions f for which all $v \in V$ are linear structures.

It follows from the Möbius inversion formula (which can be thought of as generalizing the inclusion-exclusion principle), applied to the lattice of vector subspaces of a vector space, that

$$|\mathcal{LS}^n| \;=\; \sum_{k=1}^{n} 2^{\frac{k(k-1)}{2}} \sum_{\substack{V \subseteq Z_2^n \\ \dim(V)=k}} |P(V)| \cdot (-1)^{k+1}. \tag{8}$$

Details can be found in Stanley's excellent book [30, p.116-7]. The coefficients $2^{k(k-1)/2}$ are certain values of the Möbius function for this lattice, and are calculated in Stanley's book [30, p.126-7]. The number of subspaces of dimension k is precisely $\prod_{i=0}^{k-1}(2^n - 2^i)/\prod_{i=0}^{k-1}(2^k - 2^i)$, so we have

$$2^{\frac{k(k-1)}{2}} \sum_{\substack{V \subseteq Z_2^n \\ \dim(V)=k}} |P(V)| \cdot (-1)^{k+1} = 2^{2^{n-k} + \frac{k(k+1)}{2}} \cdot \frac{\prod_{i=0}^{k-1}(2^n - 2^i)}{\prod_{i=0}^{k-1}(2^k - 2^i)}. \tag{9}$$

which is strictly decreasing as k increases when n is sufficiently large. Therefore, the expansion for $|\mathcal{LS}^n|$ in equation (8) is dominated by its first term as n becomes large. We will bound $|\mathcal{LS}^n|$ by determining the first two terms of its inclusion-exclusion expansion. Using Corollary 3.1 it follows that

$$\sum_{\substack{V \subseteq Z_2^n \\ \dim(V)=1}} |P(V)| \;=\; (2^n - 1) \cdot 2^{2^{n-1}+1} \;\stackrel{\text{def}}{=}\; U_n$$

$$2 \cdot \sum_{\substack{V \subseteq Z_2^n \\ \dim(V)=2}} |P(V)| \;=\; \frac{(2^n - 1)(2^n - 2)}{3} \cdot 2^{2^{n-2}+2} \;\stackrel{\text{def}}{=}\; L_n.$$

We have $U_n - L_n < |\mathcal{LS}^n| < U_n$, and therefore

$$1 - \frac{2^{n+1}}{3 \cdot 2^{2^{n-2}}} \;<\; \frac{|\mathcal{LS}^n|}{U_n} \;<\; 1. \tag{10}$$

It follows that $|\mathcal{LS}^n|/U_n = 1 + o(1)$, completing the proof of the theorem. $\square$

For $n = 1, 2, 3, 4$ we have that $|\mathcal{LS}^n|$ is 4, 8, 128 and 4992, respectively [24]. For larger n the bounds L_n and U_n from Theorem 3.1 may be used to yield accurate estimates of $|\mathcal{LS}^n|$. Table 2 shows bounds on $|\mathcal{LS}^n|$, and it is clear that U_n is a good estimate of the number of linear structures for $n \geq 6$.

n	U_n	L_n	L_n/U_n
5	4063232	317440	0.078125
6	$.54117 \times 10^{12}$	$.34131 \times 10^{9}$	0.63070×10^{-4}
7	$.46855 \times 10^{22}$	$.91637 \times 10^{11}$	0.19558×10^{-9}
8	$.17354 \times 10^{42}$	$.15931 \times 10^{25}$	0.91798×10^{-16}

Table 2: Bounds on $|\mathcal{LS}^n|$ for $n = 5, 6, 7, 8$.

4 Conclusion

In this paper we have provided a probabilistic analysis of algebraic nonlinearity. We have shown that a boolean function is expected to have a large algebraic nonlinearity, even when a significant number of the input variables are held constant. Linear structures characterize a property that is common to both linear and certain nonlinear functions, namely, invariance under translation of the input. It is known that if a function f has a linear structure then f can be mapped to a function for which a certain subset of the input variables may only occur in terms bounded by order 1 (partial linearity). We may enquire if there is any value in considering higher order structures, or functions that have linear transformations which map a certain set of the input variables into terms bounded by order k, $1 \leq k \leq n$. For example, an n-bit function f would have a quadratic structure if it could be mapped linearly to a function g for which there exist a subset of the variables $Y = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$, $1 \leq k \leq n$, such that

$$f(X) \;=\; g(x'_1, x'_2, \ldots, x'_{n-k}) \;+ \sum_{1 \leq j \leq k} m_j x_{i_j} + \sum_{1 \leq i < j \leq k} m_{ij} x_i x_j$$

where $\{x'_1, x'_2 \ldots, x'_{n-k}\} = \{x_1, x_2, \ldots, x_n\} - Y$, $m_j, m_{ij} \in Z_2$, $1 \le i < j \le k$. The cryptanalyst can take advantage of functions in this form if there is degeneracy in the equations. Otherwise it is known that solving equations over Z_2 with degree bounded by 2 is NP-hard [8].

References

[1] C. Adams and S. Tavares. The structured design of cryptographically good S-boxes. *Journal of Cryptology*, 3(1):27–41, 1990.

[2] M. Beale and M. F. Monaghan. Encryption using random boolean functions. In H. J. Beker and F. C. Piper, editors, *Cryptography and Coding*, pages 219–230. Clarendon Press, 1989.

[3] E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.

[4] D. Chaum and J. H. Evertse. Cryptanalysis of DES with a reduced number of rounds. *Advances in Cryptology, CRYPTO 85, H. C. Williams ed., Lecture Notes in Computer Science, vol. 218, Springer-Verlag*, pages 192–211, 1986.

[5] J. H. Evertse. Linear structures in blockciphers. *Advances in Cryptology, EUROCRYPT 87, Lecture Notes in Computer Science, vol. 304, D. Chaum and W. L. Price eds., Springer-Verlag*, pages 249–266, 1988.

[6] R. Forré. Methods and instruments for designing S-boxes. *Journal of Cryptology*, 2(3):115–130, 1990.

[7] R. Forré. The strict avalanche criterion: spectral properties of booleans functions and an extended definition. *Advances in Cryptology, CRYPTO 88, Lecture Notes in Computer Science, vol. 403, S. Goldwasser ed., Springer-Verlag*, pages 450–468, 1990.

[8] M. R. Garey and D. S. Johnson. *Computers and Intractability, A Guide to the Theory of NP-completeness.* W. H. Freeman and Co., San Francisco, 1979.

[9] M. Hall. *Combinatorial Theory.* Blaisdell Publishing Company, 1967.

[10] M. A. Harrison. *Introduction to Switching and Automata Theory.* McGraw-Hill, Inc., 1965.

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

[12] D. E. Knuth. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms.* Addison-Wesley, Reading, Ma., 1973.

[13] X. Lai. Linear structures of functions over prime fields. unpublished manuscript, 1990.

[14] W. Meier and O. Staffelbach. Nonlinear criteria for cryptographic functions. *Advances in Cryptology, EUROCRYPT 89, Lecture Notes in Computer Science, vol. 434, J. J. Quisquater, J. Vandewalle eds., Springer-Verlag*, pages 549–562, 1990.

[15] F. Mileto and G. Putzolu. Average values of quantities appearing in boolean function minimization. *IEEE Transactions on Electronic Computers*, EC-13:87–92, 1964.

[16] F. Mileto and G. Putzolu. Statistical complexity of algorithms for boolean function minimization. *Journal of the ACM*, 12(3):364–375, 1965.

[17] C. Mitchell. Enumerating boolean functions of cryptographic significance. *Journal of Cryptology*, 2(3):155–170, 1990.

[18] K. Nyberg. Perfect nonlinear S-boxes. *Advances in Cryptology, EUROCRYPT 91, Lecture Notes in Computer Science, vol. 547, D. W. Davies ed., Springer-Verlag*, pages 378–386, 1991.

[19] L. O'Connor. Enumerating nondegenerate permutations. *Advances in Cryptology, EUROCRYPT 91, Lecture Notes in Computer Science, vol. 547, D. W. Davies ed., Springer-Verlag*, pages 368–377, 1991.

[20] J. Pieprzyk and G. Finkelstein. Towards effective nonlinear cryptosystem design. *IEE proceedings*, 135, part E(6):325–335, 1988.

[21] B. Preneel, W. Van Leekwijck, L. Van Linden, R. Govaerts, and J. Vandewalle. Propogation characteristics of boolean functions. *Advances in Cryptology, EUROCRYPT 90, Lecture Notes in Computer Science, vol. 473, I. B. Damgård ed., Springer-Verlag*, pages 161–173, 1991.

[22] J. A. Reeds and J. L. Manferdelli. DES has no per round linear factors. *Advances in Cryptology, CRYPTO 84, Lecture Notes in Computer Science, vol. 196, G. R. Blakely, D. Chaum eds., Springer-Verlag*, pages 377–389, 1985.

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

[24] R. Rueppel. Private communication, March 1992.

[25] R. Rueppel. Stream ciphers. In G. Simmons, editor, *Contemporary Cryptology: the Science of Information Integrity*, pages 64–134. IEEE Press, 1991.

[26] R. A. Rueppel. *Design and Analysis of Stream Ciphers*. Springer–Verlag, 1986.

[27] J. Savage. *The Complexity of Computing*. John Wiley, 1976.

[28] T. Seigenthaler. Correlation-immunity of nonlinear combining functions for cryptographic applications. *IEEE Transactions on Information Theory*, 30(5):776–779, 1984.

[29] C. E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656–175, 1949.

[30] R. P. Stanley. *Enumerative Combinatorics, Vol. 1.* Wadsworth & Brooks/Cole, 1986.

[31] A. F. Webster and S. E. Tavares. On the design of S-boxes. *Advances in Cryptology, CRYPTO 85, H. C. Williams ed., Lecture Notes in Computer Science, vol. 218, Springer-Verlag*, pages 523–534, 1986.