A Short Proof of the Non-Existence of Certain Cryptographic Functions

K. Gopalakrishnan
Department of Computer Science
Wichita State University
Wichita KS 67260

D. R. Stinson
Department of Computer Science and Engineering and
Center for Communication and Information Science
University of Nebraska - Lincoln
Lincoln NE 68588

ABSTRACT. Several criteria have been proposed as desirable for binary cryptographic functions. Three important ones are balance, correlation-immunity and higher order strict avalanche criterion. Lloyd [7] has shown that there are no balanced, uncorrelated functions which satisfy the strict avalanche criterion of order $n-2$. In this note we give a short proof of this result using elementary combinatorial arguments. The proof relies on the solution of a recurrence relation that seems to be of interest in its own right.

1 Introduction

In this note, we will consider only functions of the form $f : [GF(2)]^n \longrightarrow GF(2)$. Several criteria have been proposed in the literature as desirable for such cryptographic functions. Three important ones are balance, correlation-immunity and higher order strict avalanche criterion. In this section we shall define these three properties.

A function is said to be *balanced* if, when all input vectors are equally likely, the output is equally likely to be 0 or 1. In other words, f is balanced if and only if

$$\sum_{x \in [GF(2)]^n} f(x) = 2^{n-1}.$$

JCMCC 20 (1996), pp. 129-137

This is an important property for almost any type of cryptographic function.

A function is said to be *correlation-immune* of mth order if knowledge of any m bits of the input vector does not give the adversary any advantage in predicting the output bit. The property of correlation-immunity is important in stream-ciphers, since combining functions which are not correlation-immune are susceptible to ciphertext-only attacks. Correlation-immunity is also desirable in the construction of S-boxes. Correlation-immune functions were defined by Siegenthaler in [10] and further studied in [4], [9], [1] and [3]. In this note, we will only be considering first order correlation-immunity.

Lemma 1.1 *Let $f : [GF(2)]^n \longrightarrow GF(2)$ be a function. Then f is balanced and correlation-immune if and only if for every i, $1 \le i \le n$, and for every $z \in GF(2)$, we have*

$$\sum_{\{x \in [GF(2)]^n:\ x_i = z\}} f(x) = 2^{n-2}.$$

Proof: Immediate. □

A function is said to satisfy the *strict avalanche criterion* (SAC), if the output bit changes with probability one half whenever a single input bit is complemented. In other words, f satisfies the SAC if and only if for every i, $1 \le i \le n$, we have

$$\sum_{x \in [GF(2)]^n} (f(x) + f(x \oplus c_i) \bmod 2) = 2^{n-1},$$

where $\oplus$ denotes bitwise addition in $GF(2)$ and c_i is the vector of length n with a 1 in the ith position and 0 elsewhere. The strict avalanche criterion was introduced by Webster and Tavares [11] in connection with the study of design of S-boxes.

The notion of strict avalanche criterion was extended by Forre [2] to consider subfunctions obtained from the original function by keeping one or more bits constant. This is also important cryptographically because, in a chosen plaintext attack, the cryptanalyst could arrange for certain input bits to be kept constant. Forre defined strict avalanche criterion of order k, where $0 \le k \le n-2$ as follows: A function $f : [GF(2)]^n \longrightarrow GF(2)$ satisfies the SAC of order k, where $1 \le k \le n-2$, if and only if any function obtained from f by keeping k of its input bits constant satisfies the SAC (for any choice of the positions and of the values of constant bits).

Lloyd has shown [7] that there are no balanced, correlation-immune functions that satisfy the strict avalanche criterion of order $n-2$. In this note, we shall prove this result in a simple manner using elementary combinatorial arguments. The proof relies on the solution of a recurrence relation that seems to be of interest in its own right.

2 Proof of Non-existence

Lloyd [6] has characterized the functions that satisfy the SAC of order $n-2$. The *algebraic normal form* (ANF) of a function is merely the expression of the function in $GF(2)$ sum-of-products form. An elegant version of the same characterization, in terms of the algebraic normal form of the function, is given in [7, p. 226]. We record this version as the following theorem.

Theorem 2.1 *Let $f : [GF(2)]^n \longrightarrow GF(2)$, where $n \geq 2$. Then f satisfies the SAC of order $n-2$ if and only if*

$$f(x) = \left(a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n + \sum_{1 \leq i < j \leq n} x_ix_j \right) \bmod 2 \quad (1)$$

for some $a_0, a_1, \ldots, a_n \in GF(2)$.

We now proceed to simplify the ANF without any loss of generality. It is easy to observe that a function f possesses all the three properties if and only if the function g defined by $g(x) = 1 + f(x)$ satisfies all the three properties. Hence, without loss of generality, we may assume that $a_0 = 0$. Further, a function f possesses all the three properties if and only if for every permutation $\pi : \{1, 2, \ldots, n\} \longrightarrow \{1, 2, \ldots, n\}$, the function g defined by

$$g(x_1, x_2, \ldots, x_n) = f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}),$$

also has all the three properties. Thus reordering the input variables does not affect the properties. Suppose exactly r of the coefficients $a_1, a_2, \ldots, a_n$ are ones and the rest are zeroes. By appropriately renaming the variables the algebraic normal form of equation (1) reduces to

$$f(x) = \left(x_1 + x_2 + \ldots + x_r + \sum_{1 \leq i < j \leq n} x_ix_j \right) \bmod 2 \quad (2)$$

for some r, $0 \leq r \leq n$.

Let $S_{n,r}$ denote the number of vectors $x \in [GF(2)]^n$ such that $f(x) = 0$. That is,

$$S_{n,r} = \left| \left\{ x \in [GF(2)]^n : x_1 + x_2 + \ldots + x_r + \sum_{1 \le i < j \le n} x_i x_j \equiv 0 \bmod 2 \right\} \right|.$$

Since f is both balanced and correlation-immune, from Lemma 1.1 we infer that for every i, $1 \le i \le n$, and for every $z \in GF(2)$,

$$\sum_{\{x \in [GF(2)]^n :\ x_i = z\}} f(x) = 2^{n-2}.$$

This condition can be expressed equivalently by the following two conditions. For every i, $1 \le i \le n$, it must be the case that

$$\sum_{\{x \in [GF(2)]^n :\ x_i = 0\}} f(x) = 2^{n-2} \quad (3)$$

$$\sum_{x \in [GF(2)]^n} f(x) = 2^{n-1} \quad (4)$$

It is easy to observe from the algebraic normal form (2) of f, that there are only two "types" of variables. That is, it is sufficient to consider only the two cases $i = 1$ and $i = r + 1$ in condition (3) instead of every i, $1 \le i \le n$. Setting $i = 1$ in condition (3) yields

$$S_{n-1,r-1} = 2^{n-2},$$

and setting $i = r + 1$ in condition (3) yields

$$S_{n-1,r} = 2^{n-2}.$$

Note that condition (4) can be expressed as

$$S_{n,r} = 2^{n-1}.$$

We summarize the above discussion as the following lemma.

Lemma 2.2 *There exists a function $f : [GF(2)]^n \longrightarrow GF(2)$, which is balanced, correlation-immune and satisfies the SAC of order $n-2$, if and only if the following three conditions are met simultaneously for some r, $0 \le r \le n$.*

$$\begin{aligned} S_{n-1,r} &= 2^{n-2} \\ S_{n-1,r-1} &= 2^{n-2} \\ S_{n,r} &= 2^{n-1} \end{aligned}$$

We now proceed to derive a recurrence relation for $S_{n,r}$.

Any vector $x \in [GF(2)]^n$ has either $x_{r+1} = 0$ or $x_{r+1} = 1$. Suppose $x_{r+1} = 0$. Then the function f reduces to a function $g : [GF(2)]^{n-1} \longrightarrow GF(2)$ whose algebraic normal form is given by

$$g(x) = \left(x_1 + x_2 + \ldots + x_r + \sum_{1 \le i < j \le n,\ i,j \ne r+1} x_i x_j \right) \bmod 2.$$

The number of vectors $x \in [GF(2)]^{n-1}$ such that $g(x) = 0$ is precisely $S_{n-1,r}$. Now, suppose that $x_{r+1} = 1$. Then the algebraic normal form of the induced function $g : [GF(2)]^{n-1} \longrightarrow GF(2)$ is

$$\begin{aligned} g(x) &= \left(x_1 + x_2 + \ldots + x_r + \sum_{1 \le i \le n,\ i \ne r+1} x_i + \sum_{1 \le i < j \le n,\ i,j \ne r+1} x_i x_j \right) \bmod 2, \\ &= \left(x_{r+2} + x_{r+3} + \ldots + x_n + \sum_{1 \le i < j \le n,\ i,j \ne r+1} x_i x_j \right) \bmod 2, \end{aligned}$$

since the arithmetic is in $GF(2)$. The number of vectors $x \in [GF(2)]^{n-1}$ such that $g(x) = 0$ is $S_{n-1,n-r-1}$. Thus we have

$$S_{n,r} = S_{n-1,r} + S_{n-1,n-r-1}. \tag{5}$$

Let us now evaluate $S_{n-1,n-r-1}$, using the recurrence relation (5):

$$\begin{aligned} S_{n-1,n-r-1} &= S_{n-2,n-r-1} + S_{n-2,(n-1)-(n-r-1)-1} \\ &= S_{n-2,n-r-1} + S_{n-2,r-1} \\ &= S_{n-2,r-1} + S_{n-2,(n-1)-(r-1)-1} \\ &= S_{n-1,r-1}. \end{aligned}$$

Substituting the above equation back in the recurrence relation (5), we obtain the following:

$$S_{n,r} = S_{n-1,r} + S_{n-1,r-1}. \tag{6}$$

It is interesting to observe that this is the same recurrence relation satisfied by the binomial coefficients (viz. Pascal's identity).

We now derive expressions for the boundary conditions $S_{n,0}$ and $S_{n,n}$. When $r = 0$, the algebraic normal form (2) reduces to

$$f(x) = \left(\sum_{1 \le i < j \le n} x_i x_j \right) \bmod 2. \tag{7}$$

The *Hamming weight* of a vector x is simply the number of positions in which 1 occurs. Note that equation (7) is symmetric in the n input bits and hence the value of $f(x)$ depends only on the Hamming weight of x. It is also trivial to observe that if x has Hamming weight k, then

$$f(x) = \binom{k}{2} \bmod 2.$$

But, $\binom{k}{2} \equiv 0 \bmod 2$ if and only if $k \equiv 0,1 \mod 4$. Thus we have

$$S_{n,0} = \sum_{k \equiv 0,1 \bmod 4,\ 0 \le k \le n} \binom{n}{k}. \tag{8}$$

When $r = n$, the algebraic normal form (2) reduces to

$$f(x) = \left(x_1 + x_2 + \ldots + x_n + \sum_{1 \le i < j \le n} x_i x_j \right) \bmod 2. \tag{9}$$

In this case again, equation (9) is symmetric in the n input bits and hence the value of $f(x)$ depends only on the Hamming weight of x. If x has Hamming weight k, then it follows that

$$f(x) = \left(k + \binom{k}{2}\right) \bmod 2.$$

Simple arithmetic shows that $\left(k + \binom{k}{2}\right) \equiv 0 \bmod 2$ if and only if $k \equiv 0,3 \mod 4$. Thus we have

$$S_{n,n} = \sum_{k \equiv 0,3 \bmod 4,\ 0 \le k \le n} \binom{n}{k}. \tag{10}$$

The recurrence relation (6), along with the boundary conditions (8) and (10), completely describes $S_{n,r}$ for $n \ge 1$ and $0 \le r \le n$. We will now derive an explicit formula for $S_{n,r}$.

First we will need the following well-known lemma which is actually a special case of a general theorem proved by C. Ramus as early as 1834 [5, p. 70, Problem 38].

Lemma 2.3

$$\begin{aligned}
\sum_{k \equiv 0 \bmod 4,\ 0 \le k \le n} \binom{n}{k} &= 2^{n-2} + 2^{\frac{n-2}{2}} \cos \frac{n\pi}{4} \\
\sum_{k \equiv 1 \bmod 4,\ 0 \le k \le n} \binom{n}{k} &= 2^{n-2} + 2^{\frac{n-2}{2}} \sin \frac{n\pi}{4} \\
\sum_{k \equiv 3 \bmod 4,\ 0 \le k \le n} \binom{n}{k} &= 2^{n-2} - 2^{\frac{n-2}{2}} \sin \frac{n\pi}{4}
\end{aligned}$$

From Lemma 2.3, the two conditions given by equations (8) and (10) become the following:

$$S_{n,0} = 2^{n-1} + 2^{\frac{n-2}{2}} \left(\cos\frac{n\pi}{4} + \sin\frac{n\pi}{4}\right) \tag{11}$$

$$S_{n,n} = 2^{n-1} + 2^{\frac{n-2}{2}} \left(\cos\frac{n\pi}{4} - \sin\frac{n\pi}{4}\right) \tag{12}$$

The next theorem gives an explicit formula for $S_{n,r}$.

Theorem 2.4

$$S_{n,r} = 2^{n-1} - 2^{\frac{n-1}{2}} \sin\left[\left(r + \frac{7n-1}{2}\right)\frac{\pi}{2}\right], \tag{13}$$

for all $n \geq 1$ *and for all* $0 \leq r \leq n$.

Proof: When $r = 0$, equation (13) is the same as equation (11); and when $r = n$, equation (13) is the same as equation (12) by basic trigonometric identities. It is also a routine matter to verify that the formula given in Theorem 2.4 satisfies the recurrence relation (6). □

We shall now state and prove the main theorem.

Theorem 2.5 *There are no functions* $f : [GF(2)]^n \longrightarrow GF(2)$, $n \geq 2$, *which are balanced, correlation-immune and satisfy the strict avalanche criterion of order* $n-2$.

Proof: Suppose $f : [GF(2)]^n \longrightarrow GF(2)$ is a function which satisfies all the three abovementioned properties. Then from Lemma 2.2 it follows that there exists an r, $0 \leq r \leq n$, which satisfies the following three conditions simultaneously.

$$S_{n-1,r} = 2^{n-2} \tag{14}$$

$$S_{n-1,r-1} = 2^{n-2} \tag{15}$$

$$S_{n,r} = 2^{n-1} \tag{16}$$

Actually, in view of the recurrence relation (6), condition (16) is redundant. Thus the function f possesses all the three properties if and only if conditions (14) and (15) are met simultaneously for some r, $0 \leq r \leq n$.

Let

$$\alpha = \left(r + \frac{7n-8}{2}\right)\frac{\pi}{2}. \tag{17}$$

Using the explicit formula provided by Theorem 2.4 and the notation (17), we express the conditions (14) and (15) by the following equations:

$$2^{n-2} - 2^{\frac{n-2}{2}} \sin \alpha \;=\; 2^{n-2} \tag{18}$$

$$2^{n-2} - 2^{\frac{n-2}{2}} \sin\left(\alpha - \frac{\pi}{2}\right) \;=\; 2^{n-2} \tag{19}$$

Clearly conditions (18) and (19) can be simultaneously satisfied if and only if

$$\sin \alpha = \sin\left(\alpha - \frac{\pi}{2}\right) = 0.$$

However, this is obviously impossible and hence the theorem is proved. □

3 Remarks

Our Theorem 2.5 can also be obtained as a corollary of Lloyd's Proposition 3.8 [7]. As well, an anonymous referee has pointed out that yet another approach to proving the result of this paper is to use tools developed in [8, Chapter 15] on properties of quadratic boolean functions.

Acknowledgments

The authors' research was supported by NSF grant CCR-9121051.

References

[1] P. Camion, C. Carlet, P. Charpin, and N. Sendrier. On correlation-immune functions. In *Advances in Cryptology – CRYPTO '91*, pages 86–100. Springer-Verlag, 1992.

[2] R. Forre. The Strict Avalanche Criterion: Spectral Properties of Boolean Functions and an Extended Definition. In *Advances in Cryptology – CRYPTO '88*, pages 450–468. Springer-Verlag, 1990.

[3] K. Gopalakrishnan and D. R. Stinson. Three Characterizations of Non-binary Correlation-Immune and Resilient Functions. To appear in *Designs, Codes and Cryptography.*

[4] X. Guo-Zhen and J. L. Massey. A Spectral Characterization of Correlation-Immune Combining Functions. *IEEE Trans. Inform. Theory*, 34 (1988), 569–571.

[5] D. E. Knuth. *Fundamental Algorithms, Second Edition.* The Art of Computer Programming, vol. 1. Addison Wesley, 1973.

[6] S. LLOYD. Counting functions satisfying a higher order strict avalanche criterion. In *Advances in Cryptology – EUROCRYPT '89*, pages 63–74. Springer-Verlag, 1990.

[7] S. LLOYD. Balance, uncorrelatedness and the strict avalanche criterion. *Discrete Applied Mathematics*, 41 (1993), 223–233.

[8] F. J. MACWILLIAMS AND N. J. A. SLOANE. *The Theory of Error-Correcting Codes.* North-Holland, 1977.

[9] R. A. RUEPPEL. *Analysis and Design of Stream Ciphers.* Springer-Verlag, 1986.

[10] T. SIEGENTHALER. Correlation immunity of nonlinear combining functions for cryptographic applications. *IEEE Trans. Inform. Theory*, 30 (1984), 776–780.

[11] A. F. WEBSTER AND S. E. TAVARES. On the design of S-boxes. In *Advances in Cryptology – CRYPTO '85*, pages 523–534. Springer-Verlag, 1986.