Correspondence

A Spectral Characterization of Correlation-Immune Combining Functions

XIAO GUO-ZHEN AND JAMES L. MASSEY, FELLOW, IEEE

Abstract—**It is shown that a Boolean combining function $f(x)$ of n variables is mth-order correlation immune if and only if its Walsh transform $F(\omega)$ vanishes for all ω with Hamming weights between 1 and m, inclusive. This result is used to extend slightly Siegenthaler's characterization of the algebraic normal form of correlation-immune combining functions.**

I. Introduction

A common form of running key generator for use in stream ciphers consists of n binary linear feedback shift registers (LFSR's) whose outputs are combined by a memoryless device, as illustrated in Fig. 1. The Boolean combining function $f(x_1, x_2, \cdots, x_n)$ generally is chosen to be nonlinear over the finite field GF(2) so as to increase the "linear complexity" of the resulting running key stream, i.e., to increase the length of the shortest LFSR that can generate this binary sequence.

Siegenthaler [1] has recently shown that several combining functions previously proposed in the literature can be broken by a ciphertext-only correlation attack. In subsequent work [2] Siegenthaler introduced the concept of mth-order correlation immunity for combining functions as a measure of their resistance against such correlation attacks. He also showed how, by iteration, to construct a limited family of mth-order correlation-immune combining functions for every m, $1 \le m \le n$.

In this correspondence we characterize all mth-order correlation-immune combining functions for every m, $1 \le m \le n$, in terms of their Walsh transforms. We use this result to extend slightly Siegenthaler's characterization of the algebraic normal form of the combining function of correlation-immune combining functions.

II. Walsh Transform of a Boolean Function

Let $x = (x_1, x_2, \cdots, x_n)$ and $\omega = (\omega_1, \omega_2, \cdots, \omega_n)$ be n-tuples over GF(2), and define their *dot product* as

$$x \cdot \omega = x_1\omega_1 + x_2\omega_2 + \cdots + x_n\omega_n.$$

Let $f(x)$ be any real-valued function whose domain is the vector space GF(2)^n of binary n-tuples. Then the *Walsh transform* [3] of $f(x)$ is another real-valued function over GF(2)^n that can be defined as

$$F(\omega) = \sum_x f(x)(-1)^{x \cdot \omega} \tag{1}$$

where (here and hereafter) the sum is over all x in GF(2)^n and $x \cdot \omega$ in the exponent is treated as the integer 0 or 1 rather than as an element of GF(2). The function $f(x)$ can be recovered by the inverse Walsh transform

$$f(x) = 2^{-n} \sum_\omega F(\omega)(-1)^{x \cdot \omega}. \tag{2}$$

Manuscript received August 14, 1985; revised December 22, 1986. This correspondence was presented in part at the IEEE International Symposium on Information Theory, Brighton, England, June 24–28, 1985.

Xiao Guo-Zhen is with the Department of Mathematics, Northwest Telecommunication Engineering Institute, Xian, China.

J. L. Massey is with the Institute for Signal and Information Processing, Swiss Federal Institute of Technology, ETH-Zentrum, CH-8092 Zurich, Switzerland.

IEEE Log Number 8821587.

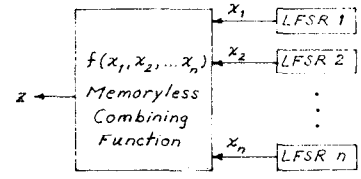

Fig. 1. Common form of running key generator.

By treating the values of a Boolean function as the real numbers 0 and 1, one can use (1) to define the Walsh transform $F(\omega)$ of a Boolean function of n variables [4], and one can then use (2) to recover $f(x)$ from its Walsh transform. Hereafter, $f(x)$ will always denote a Boolean function of n binary variables.

It is convenient in Walsh transforming of Boolean functions to introduce the integer-valued functions

$$N_{f0}(\omega) = \#\{x: f(x) = 1 \text{ and } x \cdot \omega = 0\} \tag{3a}$$

$$N_{f1}(\omega) = \#\{x: f(x) = 1 \text{ and } x \cdot \omega = 1\} \tag{3b}$$

where $\#\{\cdot\}$ denotes the cardinality of the indicated set. It follows from (1) that

$$F(\omega) = N_{f0}(\omega) - N_{f1}(\omega). \tag{4}$$

III. Correlation Immunity and Walsh Transforms

A binary [i.e., GF(2)-valued] random variable is said to be *balanced* if it is equally likely to take on the values 0 and 1. Siegenthaler [2] has defined the combining function $f(x)$ to be *mth-order correlation immune* if the random variable $Z = f(X_1, X_2, \cdots, X_m)$ is statistically independent of every set of m random variables chosen from the balanced and independent binary random variables $X_1, X_2, \cdots, X_N$. To see the implications of correlation immunity on the Walsh transform of $f(x)$, we require the following result that we prove for an arbitrary discrete random variable Z.

Lemma: The discrete random variable Z is independent of the m independent binary random variables $Y_1, Y_2, \cdots, Y_m$ if and only if Z is independent of the sum $\lambda_1 Y_1 + \lambda_2 Y_2 + \cdots + \lambda_m Y_m$ for every choice of $\lambda_1, \lambda_2, \cdots, \lambda_m$, not all zeros, in GF(2).

Remark: In our subsequent use of this lemma the random variables $Y_1, Y_2, \cdots, Y_m$ will all be balanced; however, the proof of the more general result is as easy as that of this special case.

Proof: The necessity of the stated condition is obvious, and it remains only to show its sufficiency. Sufficiency is trivial for $m = 1$. Consider $m = 2$, and suppose that Z is independent of Y_1, of Y_2, and of $Y_1 + Y_2$. This implies the following relations among probability distributions for every possible value z of Z with $P_Z(z) > 0$:

$$\begin{aligned}
&P_{Y_1Y_2|Z}(1,1|z) + P_{Y_1Y_2|Z}(1,0|z) = P_{Y_1|Z}(1|z) = p_1 \\
&P_{Y_1Y_2|Z}(1,1|z) + P_{Y_1Y_2|Z}(0,1|z) = P_{Y_2|Z}(1|z) = p_2 \\
&P_{Y_1Y_2|Z}(1,0|z) + P_{Y_1Y_2|Z}(0,1|z) = P_{Y_1+Y_2|Z}(1|z) \\
&\qquad\qquad = p_1(1-p_2) + (1-p_1)p_2 \\
&P_{Y_1Y_2|Z}(1,0|z) + P_{Y_1Y_2|Z}(0,0|z) = P_{Y_2|Z}(0|z) = 1 - p_2
\end{aligned} \tag{5}$$

where $p_1 = P_{Y_1}(1)$ and $p_2 = P_{Y_2}(1)$. It is easily checked that these

four linear equations are independent and have the unique solution

$$P_{Y_1Y_2|Z}(0,1|z) = (1-p_1)p_2$$
$$P_{Y_1Y_2|Z}(1,0|z) = p_1(1-p_2)$$
$$P_{Y_1Y_2|Z}(0,0|z) = (1-p_1)(1-p_2)$$
$$P_{Y_1Y_2|Z}(1,1|z) = p_1p_2 \tag{6}$$

which is just the product distribution for Y_1 and Y_2. Thus the pair Y_1, Y_2 is independent of Z, and hence the triple Y_1, Y_2, Z is independent.

Now assume that the sufficiency of the condition stated in the lemma has been established for $m < M$. To show sufficiency for $m = M$, we argue as follows. Suppose that Z is independent of the sum $\lambda_1Y_1 + \lambda_2Y_2 + \cdots + \lambda_MY_M$ for all $\lambda_1, \lambda_2, \cdots, \lambda_M$, not all zeroes, in GF(2). Choosing $\lambda_1 \equiv \lambda_2 \equiv 0$ implies, by the induction hypothesis, that $Y_3, \cdots, Y_M, Z$ are independent. Moreover, choosing $\lambda_2 \equiv 0$ implies, by the induction hypothesis, that Y_1 is independent of $Y_3, \cdots, Y_M, Z$. Similarly, the choice $\lambda_1 = 0$ and the choice $\lambda_1 \equiv \lambda_2$ imply, respectively, that Y_2 is independent of $Y_3, \cdots, Y_M, Z$ and that the binary random variable $Y_1 + Y_2$ is independent of $Y_3, \cdots, Y_M, Z$. Thus for every choice of $y_3, \cdots, y_M, z$, (5) and (6) still apply after replacing

$$P_{Y_1Y_2|Z}(y_1, y_2|z)$$

by

$$P_{Y_1Y_2|Y_3 \cdots Y_MZ}(y_1, y_2|y_3, \cdots, y_M, z)$$

everywhere. It follows from (6) so rewritten that the pair Y_1, Y_2 is independent of $Y_3, \cdots, Y_M, Z$ and thus that the random variables $Y_1, Y_2, Y_3, \cdots, Y_m, Z$ are all independent, as was to be shown.

In what follows, $X_1, X_2, \cdots, X_n$ will always denote n independent and balanced binary random variables, $\boldsymbol{X}$ will always denote the random n-tuple $[X_1, X_2, \cdots, X_n]$, and Z will always denote the binary random variable $f(\boldsymbol{X})$. Let $W(\boldsymbol{\omega})$ denote the Hamming weight of the binary n-tuple $\boldsymbol{\omega}$, i.e., the number of nonzero components in $\boldsymbol{\omega}$. Then for any $\boldsymbol{\omega} \neq \boldsymbol{0}$, the random variable $\boldsymbol{\omega} \cdot \boldsymbol{X} = \omega_1X_1 + \cdots + \omega_nX_n$ is a GF(2) sum of $W(\boldsymbol{\omega})$ of the random variables $X_1, \cdots, X_N$. Because all 2^{n-1} values $\boldsymbol{x}$ of $\boldsymbol{X}$ that give $\boldsymbol{x} \cdot \boldsymbol{\omega} = b$ are equally likely, it follows from (3a) and (3b) that

$$P_{Z|\boldsymbol{\omega}\cdot\boldsymbol{X}}(1|b) = 2^{-n+1}N_{fb}(\boldsymbol{\omega}), \qquad \text{for } \boldsymbol{\omega} \neq \boldsymbol{0} \tag{7}$$

for $b = 0,1$. The relation (7) is the key to the following theorem, which is the main result of this correspondence.

Theorem: The Boolean combining function $f(\boldsymbol{x})$ for n binary variables is mth-order correlation immune, where $1 \leq m \leq n$, if and only if its Walsh transform satisfies

$$F(\boldsymbol{\omega}) = 0, \qquad \text{for } 1 \leq W(\boldsymbol{\omega}) \leq m.$$

Proof: By definition, $f(\boldsymbol{x})$ is mth-order correlation immune if and only if Z is statistically independent of every subset of m or fewer of the random variables $X_1, X_2, \cdots, X_n$. It then follows from the Lemma that $f(\boldsymbol{x})$ is mth-order correlation immune if and only if Z is independent of every random variable $\boldsymbol{\omega} \cdot \boldsymbol{X}$ for which $1 \leq W(\boldsymbol{\omega}) \leq m$. However, (7) shows that the binary random variable Z is independent of $\boldsymbol{\omega} \cdot \boldsymbol{X}$ just when $N_{f0}(\boldsymbol{\omega}) = N_{f1}(\boldsymbol{\omega})$. Thus $f(\boldsymbol{x})$ is mth-order correlation immune if and only if $N_{f0}(\boldsymbol{\omega}) = N_{f1}(\boldsymbol{\omega})$ for $1 \leq W(\boldsymbol{\omega}) \leq m$. However, (4) shows that this necessary and sufficient condition is equivalent to $F(\boldsymbol{\omega}) = 0$ for $1 \leq W(\boldsymbol{\omega}) \leq m$, which proves the theorem.

Because the 2^n values of $\boldsymbol{X}$ are equally likely and because $F(\boldsymbol{0})$, according to (1), is just the number of $\boldsymbol{x}$ such that $f(\boldsymbol{x}) = 1$, it follows that the probability distribution for $Z = f(\boldsymbol{X})$ satisfies

$$P_Z(1) = 2^{-n}F(\boldsymbol{0}). \tag{8}$$

Moreover, it follows from (3a) and (3b) that for all $\boldsymbol{\omega}$, $N_{f0}(\boldsymbol{\omega}) + N_{f1}(\boldsymbol{\omega})$ equals the number of $\boldsymbol{x}$ such that $f(\boldsymbol{x}) = 1$, and hence that

$$N_{f0}(\boldsymbol{\omega}) + N_{f1}(\boldsymbol{\omega}) = F(\boldsymbol{0}), \qquad \text{all } \boldsymbol{\omega}.$$

Thus it follows from (4) that $F(\boldsymbol{\omega}) = 0$ is equivalent to $N_{fb}(\boldsymbol{\omega}) = F(\boldsymbol{0})/2$ for $b = 0,1$. Thus the condition of the Theorem may be written in the following equivalent form.

Corollary: The Boolean combining function $f(\boldsymbol{x})$ is mth-order correlation immune if and only if

$$N_{fb}(\boldsymbol{\omega}) = F(\boldsymbol{0})/2, \qquad \text{for } 1 \leq W(\boldsymbol{\omega}) \leq m, \; b = 0,1.$$

IV. Application to Algebraic Normal Forms

It is common in cryptology to work with combining functions expressed in algebraic normal form (i.e., in GF(2) sum-of-products form), namely,

$$f(\boldsymbol{x}) = a_0 + a_1x_1 + \cdots + a_nx_n + a_{12}x_1x_2 + \cdots + a_{12\ldots n}x_1x_2\cdots x_n. \tag{9}$$

It would thus be highly desirable to express the condition for correlation immunity in terms of the coefficients in the algebraic normal form of $f(\boldsymbol{x})$. We now give some partial results in this direction.

We begin by expressing the coefficients on the right of (9) in terms of the Walsh transform $F(\boldsymbol{\omega})$. Note that $a_0 = f(\boldsymbol{0})$ so that (2) gives

$$a_0 = 2^{-n}\sum_{\boldsymbol{\omega}} F(\boldsymbol{\omega}). \tag{10}$$

Next, let $U(i_1, i_2, \cdots, i_k)$ be the vector space of all 2^k binary n-tuples $\boldsymbol{x}$ such that $x_i = 0$ when $i \notin \{i_1, \cdots, i_k\}$. Any product of j of the variables $x_1, x_2, \cdots, x_n$ vanishes for all $\boldsymbol{x}$ in $U(i_1, i_2, \cdots, i_k)$ unless these variables all have indices in $\{i_1, \cdots, i_k\}$, in which case the product equals 1 for exactly 2^{k-j} n-tuples $\boldsymbol{x}$ in $U(i_1, \cdots, i_k)$. Thus (9) implies

$$a_{i_1i_2\cdots i_k} = \sum_{\boldsymbol{x} \in U(i_1, \cdots, i_k)} f(\boldsymbol{x}) \pmod 2 \tag{11}$$

where here and hereafter we treat the values of $f(\boldsymbol{x})$ as the real numbers 0 and 1, and we write (mod 2) after an expression only when that expression must be equal to an integer and to mean 0 or 1 according as that integer is even or odd, respectively. Substituting (2) into (11) gives

$$a_{i_1i_2\cdots i_k} = 2^{-n}\sum_{\boldsymbol{\omega}} F(\boldsymbol{\omega}) \sum_{\boldsymbol{x} \in U(i_1, \cdots, i_k)} (-1)^{\boldsymbol{x}\cdot\boldsymbol{\omega}} \pmod 2. \tag{12}$$

Now if $\boldsymbol{\omega}$ has any nonzero component with index in $\{i_1, \cdots, i_k\}$, then exactly half of the vectors $\boldsymbol{x}$ in $U(i_1, \cdots, i_k)$ will yield $\boldsymbol{x} \cdot \boldsymbol{\omega} = 1$ so that the second sum in (12) will vanish; otherwise, $\boldsymbol{x} \cdot \boldsymbol{\omega} = 0$ for every $\boldsymbol{x}$ in $U(i_1, \cdots, i_k)$ so that this sum equals 2^k. Thus defining $V(i_1, \cdots, i_k)$ as the vector space of all 2^{n-k} binary n-tuples $\boldsymbol{\omega}$ such that $\omega_i = 0$ when $i \in \{i_1, \cdots, i_k\}$, we have from (12)

$$a_{i_1i_2\cdots i_k} = 2^{-n+k} \sum_{\boldsymbol{\omega} \in V(i_1, \cdots, i_k)} F(\boldsymbol{\omega}) \pmod 2, \tag{13}$$

which is our desired expression.

To relate (13) to correlation immunity, we note first that

$$W(\boldsymbol{\omega}) \leq n - k, \qquad \text{if } \boldsymbol{\omega} \in V(i_1, \cdots, i_k). \tag{14}$$

Thus the Theorem implies that if $f(\boldsymbol{x})$ is mth-order correlation immune and $n - k \leq m$, then the only nonvanishing term in the

sum (13) will be that for $\omega = \mathbf{0}$; hence

$$a_{i_1 i_2 \cdots i_k} = 2^{-n+k} F(\mathbf{0}) \pmod 2, \qquad \text{if } k \geq n-m.$$

However, upon noting that $2^{-n+k}F(0)$ must be an integer for $k \geq n-m$ and thus must be an even integer for $k > n-m$, we see that this relation is equivalent to

$$a_{i_1 i_2 \cdots i_{n-m}} = 2^{-m} F(\mathbf{0}) \pmod 2, \qquad \text{if } n > m \tag{15}$$

and

$$a_{i_1 i_2 \cdots i_k} = 0, \qquad \text{if } k > n-m. \tag{16}$$

The necessity of (16), when $f(\boldsymbol{x})$ is mth-order correlation immune has already been shown by Siegenthaler [2, theorem 1]; the necessity of (15), i.e., that all coefficients of $(n-m)$th-order product terms must be equal, is new.

Example 1: The Walsh transform $F(\omega_1, \omega_2, \omega_3)$ of $f(x_1, x_2, x_3) = x_1 + x_2 + x_3 + x_1x_2 + x_1x_3 + x_2x_3$ satisfies $F(1,0,0) = F(0,1,0) = F(0,0,1) = 0$, but $F(1,1,0) = -2$. Thus by the Theorem, $f(\boldsymbol{x})$ is correlation immune of order $m = 1$ (but not $m = 2$). Notice that the coefficients of all product terms of order $n - m = 2$ in $f(\boldsymbol{x})$ are equal to 1.

Example 2: The Walsh transform $F(\omega_1, \omega_2, \omega_3, \omega_4)$ of $f(x_1, x_2, x_3, x_4) = x_1 + x_2 + x_3 + x_4 + x_1x_3 + x_2x_3 + x_1x_4 + x_2x_4$ satisfies $F(1,0,0,0) = F(0,1,0,0) = F(0,0,1,0) = F(0,0,0,1) = 0$, but $F(0,0,1,1) = -4$. Thus by the Theorem, $f(\boldsymbol{x})$ is correlation immune of order $m = 1$ (but not $m = 2$). Notice that the coefficients of all product terms of order $n - m = 3$ in $f(\boldsymbol{x})$ are equal to 0.

Finally, we note that one often demands in cryptographic applications that the random variable $Z = f(\boldsymbol{X})$ be balanced. From (8), we see that Z is balanced if and only if $F(\mathbf{0}) = 2^{n-1}$. In the previous examples, $F(0,0,0) = 6 \neq 2^{n-1} = 4$ and $F(0,0,0,0) = 12 \neq 2^{n-1} = 8$; thus $Z = f(X)$ is not balanced. When Z is balanced, (15) becomes

$$a_{i_1 i_2 \cdots i_{n-m}} = 2^{n-m-1} \pmod 2, \qquad \text{if } n > m$$

and hence these coefficients of order $n - m$ must all vanish if $m \neq n-1$ and must all equal 1 if $m = n-1$, as has already been noted by Siegenthaler [2].

V. Concluding Remark

Golomb [4], who was apparently the first to consider Walsh transforms of Boolean functions, used this technique at the Jet Propulsion Laboratory (JPL) in the late 1950's for, among other things, the design of an interplanetary ranging system [5]. The objective then was to design a Boolean combiner for shift-register sequences of short relatively prime periods to produce a sequence whose period was the product of the component periods and which was highly correlated with each component sequence to facilitate the calculation of range. This ranging problem is virtually dual to the cryptographic problem posed by Siegenthaler [2] and considered in this correspondence; the interested reader can find details of the ranging problem in the work of Titsworth [6]. In fact, our results for $m = 1$ (first-order correlation immunity) can be deduced immediately from the work of Golomb and Titsworth; the Lemma in this paper is needed, however, to extend the results to the general case where $m > 1$.

Acknowledgment

The authors are grateful to Thomas Siegenthaler of the Institute for Communication Technology, Swiss Federal Institute of Technology, Zurich, for his constructive criticisms of an earlier draft of this paper and for several helpful suggestions. The authors are likewise grateful to Dr. Giancarlo Duella of the Institute for Microelectronics, University of Neuchâtel, Switzerland, for pointing out that an assumption in our Lemma that $Y_1, \cdots, Y_m$ be balanced could be avoided.

References

[1] T. Siegenthaler, "Decrypting a class of stream ciphers using ciphertext only," *IEEE Trans. Comput.*, vol. C-34, pp. 81–85, Jan. 1985.
[2] ____, "Correlation immunity of nonlinear combining functions for cryptographic applications," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 776–780, Oct. 1984.
[3] M. G. Karpovsky, *Finite Orthogonal Series in the Design of Digital Devices*. New York and Jerusalem: Wiley and IUP, 1976.
[4] S. W. Golomb, "On the classification of Boolean functions," *IEEE Trans. Inform. Theory*, vol. IT-5, pp. 176–186, May 1959. (Also appears in S. W. Golomb, *Shift Register Sequences*. San Francisco, CA: Holden-Day, 1967, ch. VIII.)
[5] ____, private communication, Mar. 1986.
[6] R. C. Titsworth, "Optimal ranging codes," *IEEE Trans. Space Elec., Telem.*, vol. SET-10, pp. 19–20, Mar. 1964.

On PPM Sequences with Good Autocorrelation Properties

COSTAS N. GEORGHIADES

***Abstract*—The problem of designing sequences of Q-ary pulse-position-modulation symbols that have good periodic autocorrelation properties is investigated. Two cases are considered. In the first it is assumed that only slot synchronization is present and thus cyclic shifts are one slot at a time; in the second PPM symbol synchronization is present, in which case cyclic shifts are by one symbol (Q slots) at a time. In both cases upper bounds are derived on the maximum peak-to-sidelobe distance, which are shown through a computer search to be nearly tight. When symbol synchronization is present, the bound reduces to the Plotkin bound, but it is slightly tighter in general.**

I. Introduction and Motivation

Sequences with "good" autocorrelation properties have applications in a variety of areas, including ranging, spread spectrum, and synchronization. The motivation for our work stems from previous results on frame synchronization for the optical direct-detection channel utilizing pulse-position modulation (PPM) [1], [2]. However, the general results derived herein are by no means limited to the optical channel, although they are of current interest for the latter where PPM has been shown to be optimal in a variety of ways [3]–[5]. Under PPM, information is conveyed by the position of a signal pulse in one (and only one) of Q subintervals (slots) dividing the symbol interval. This restriction will be referred to in the sequel as the PPM constraint.

A standard approach to the frame synchronization problem is to insert periodically in the data stream a special sequence and achieve synchronization at the receiver by locating this pattern. Maximum-likelihood (ML) frame synchronizers for binary modulation already have been derived for the Gaussian channel [6] and for on–off-keying and PPM for the optical channel [1], [2]. Although the structure of the optimal synchronizers does not depend on the specific synchronization sequence used, their performance is critically dependent on the choice of a "good" sequence. In evaluating the performance of the ML synchronizer for the Gaussian channel with binary modulation, Massey [6] used the already existing Parker [7] and Neuman–Hoffman [8] sequences. A literature search by the author in [1], [2] revealed

Manuscript received October 13, 1986; revised May 12, 1987. This correspondence was presented in part at the IEEE International Symposium on Information Theory, Ann Arbor, MI, October 1986.

The author is with the Electrical Engineering Department, Texas A&M University, College Station, TX 77843.

IEEE Log Number 8821812.