On Constructions and Nonlinearity of Correlation Immune Functions *

Jennifer Seberry
Xian-Mo Zhang
Yuliang Zheng

The Centre for Computer Security Research
Department of Computer Science
The University of Wollongong
Wollongong, NSW 2522, AUSTRALIA
E-mail: {jennie,xianmo,yuliang}@cs.uow.edu.au

November 2, 1993

Abstract

A Boolean function is said to be correlation immune if its output leaks no information about its input values. Such functions have many applications in computer security practices including the construction of key stream generators from a set of shift registers. Finding methods for easy construction of correlation immune functions has been an active research area since the introduction of the notion by Siegenthaler. In this paper we study balanced correlation immune functions using the theory of Hadamard matrices. First we present a simple method for directly constructing balanced correlation immune functions of any order. Then we prove that our method generates exactly the same set of functions as that obtained using a method by Camion, Carlet, Charpin and Sendrier. Advantages of our method over Camion et al's include (1) it allows us to calculate the nonlinearity, which is a crucial criterion for cryptographically strong functions, of the functions obtained, and (2) it enables us to discuss the propagation characteristics of the functions. Two examples are given to illustrate our construction method. Finally, we investigate methods for obtaining new correlation immune functions from known correlation immune functions. These methods provide us with a new avenue towards understanding correlation immune functions.

*The first author was supported in part by the Australian Research Council under the reference numbers A49130102, A9030136, A49131885 and A49232172, the second author by A49130102, and the third author by A49232172.

Key Words

correlation immunity, stream cipher, nonlinearity, Hadamard matrix, cryptography.

1 Introduction

The main component of a stream cipher is a key stream generator which produces from a random seed a sequence of pseudo-random bits. These pseudo-random bits are added modulo 2 to bits in a plaintext and the resulting stream, a ciphertext, is sent to a receiver. The receiver can recover the plaintext by adding modulo 2 to the ciphertext the output of the stream generator with the same seed.

A common method for obtaining key stream generators is to combine a set of shift registers with a nonlinear function. Blaser and Heinzmann [1] observed that if the combining function leaks information about its component functions, then the work needed in attacking the cryptosystem can be significantly reduced. This idea was further developed by Siegenthaler in [8] where a new concept called correlation immune functions was introduced. Since then the topic has been an active research area and correlation immunity has become one of the central design criteria for stream ciphers based on shift registers [4, 5].

For practical applications, finding methods for easy construction of correlation immune functions is of most importance. In [8] Siegenthaler presented the first method for constructing (balanced) correlation immune functions. His method is recursive in nature and hence not very satisfactory in practical applications. Camion et al studied correlation immune functions from the point view of algebraic coding theory, and presented a method for constructing correlation immune functions of any order [2].

In this paper we study correlation immune functions using the theory of Hadamard matrices. First we present a method for directly constructing balanced correlation immune functions of any order. We then prove that our method generates exactly the same set of correlation immune functions as that obtained using Camion et al's method. Advantages of our method over Camion et al's include that, in addition to their orders of correlation immunity and algebraic degrees, it gives the nonlinearity and propagation characteristics of the functions obtained. We also study methods for constructing correlation immune functions on a higher dimensional space by combining known correlation immune functions on a lower dimensional space. The nonlinearity of functions thus constructed is also investigated.

The organization of the rest of the paper is as follows. Section 2 introduces notations and definitions that are needed in the paper. Section 3 reviews the previous construction methods for correlation immune functions. Our new construction method is described in Section 4. In the same section we also prove that the new construction method generates exactly the same set of correlation immune functions as that by Camion et al's method. Section 5 discusses the algebraic degree, nonlinearity and propagation characteristics of functions obtained using the new method. Two examples are shown in the same section. Section 6 is devoted to the combination of known correlation immune functions. Three combination methods are shown

in the section, among which the first one can be viewed as an extension of the new construction method described in Section 4. The paper concludes with some remarks in Section 7.

2 Preliminaries

We consider V_m, the vector space of m tuples of elements from $GF(2)$. Note that there is a natural one to one correspondence between vectors in V_m and integers in $[0, 2^m - 1]$. This allows us to order the vectors according to their corresponding integer values. For convenience, we denote by α_i the vector in V_m whose integer representation is i.

Let f be a function from V_m to $GF(2)$ (or simply a function on V_m). Since f can be expressed as a unique polynomial in m coordinates $x_1, x_2, \ldots, x_m$, we will identify f with its unique multi-variable polynomial $f(x)$ where $x = (x_1, x_2, \ldots, x_m)$. To distinguish between a vector of coordinates and an individual coordinate, the former will be strictly denoted by w, x, y or z, while the later strictly by w_i, x_i, y_i, z_i or u, where i is an index. The algebraic degree of f is defined as the number of coordinates in its longest term when it is represented in the algebraic normal form. f is called an *affine function* if it takes the form of $f(x) = a_1x_1 \oplus \cdots \oplus a_mx_m \oplus c$, where a_j, $c \in GF(2)$. In particular, f is called a *linear* function if $c = 0$.

The *sequence* of f on V_m is a $(1, -1)$-sequence defined by $((-1)^{f(\alpha_0)}, (-1)^{f(\alpha_1)}, \ldots, (-1)^{f(\alpha_{2^m-1})})$, and the *truth table* of f is a $(0\ ,\ 1)$-sequence defined by $(f(\alpha_0), f(\alpha_1), \ldots, f(\alpha_{2^m-1}))$. f is said to be *balanced* if the truth table of f has 2^{m-1} zeros (ones).

The following notation will be used in this paper. Let $\alpha = (a_1, \cdots, a_m)$ and $\beta = (b_1, \cdots, b_m)$ be two vectors (or sequences), the *scalar product* of α and β, denoted by $\langle \alpha, \beta \rangle$, is defined as the sum of the component-wise multiplications. In particular, when α and β are from V_m, $\langle \alpha, \beta \rangle = a_1b_1 \oplus \cdots \oplus a_mb_m$, where the addition and multiplication are over $GF(2)$, and when α and β are $(1, -1)$-sequences, $\langle \alpha, \beta \rangle = \sum_{i=1}^{m} a_ib_i$, where the addition and multiplication are over the reals.

Now we introduce the concept of *correlation immune functions*, the central topic treated in this paper. Let f be a function on V_m. Let X be a random variable taking on values $x \in V_m$ with uniform probability 2^{-m}, let X_i be the random variable corresponding to the ith coordinate value $x_i \in GF(2)$, and let Y be the random variable produced by the function f, i.e., $Y = f(X)$. f is said to be a *kth-order correlation immune function* if the random variable Y is statistically independent of any subset $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ of k coordinates [8].

Xiao and Massey gave an equivalent definition for correlation immunity in terms of *Walsh transformations* [3]. The Walsh transformation $\hat{f}$ of a function f on V_m is defined as the real-valued function

$$\hat{f}(\beta) = \sum_{x \in V_m} f(x)(-1)^{\langle \beta, x \rangle},$$

where $\beta \in V_m$. Note that in the sum, $f(x)$ and $\langle \beta, x \rangle$ are regarded as real-valued functions.

Definition 1 *Let f be a function on V_m. f is a* kth-order correlation immune function *if its Walsh transformation satisfies $\hat{f}(\beta) = 0$ for all $\beta \in V_m$ with $1 \leq W(\beta) \leq k$, where $W(\beta)$ indicates the Hamming weight of, i.e., the number of the nonzero components in, a vector β.*

A relevant topic, correlation immune functions with memory, was studied in [4]. The next lemma is useful for constructing correlation immune functions with a view to using Hadamard matrices.

Lemma 1 *Let g be a function on V_m and let η be its sequence. Also let $x = (x_1, x_2, \ldots, x_m)$. Then g is a* kth-order correlation immune function *if and only if $\langle \eta, \ell \rangle = 0$ for any ℓ, where ℓ is the sequence of a linear function $h(x) = \langle \alpha, x \rangle$ on V_m constrained by $1 \leqq W(\alpha) \leqq k$.*

Proof. Note that

$$\begin{aligned} \langle \eta, \ell \rangle &= \sum_{x \in V_m} (-1)^{g(x)} (-1)^{h(x)} = \sum_{x \in V_m} (-1)^{g(x) + \langle \alpha, x \rangle} \\ &= \sum_{x \in V_m} (-1)^{\langle \alpha, x \rangle} - 2 \sum_{x \in V_m} g(x) (-1)^{\langle \alpha, x \rangle} \\ &= -2\hat{g}(\alpha). \end{aligned}$$

Thus $\langle \eta, \ell \rangle = 0$ if and only if $\hat{g}(\alpha) = 0$ (See also Section 4.2, [2]). □

The order k of correlation immunity of a function on V_m and its algebraic degree d are constrained by the relation $k + d \leq m$. The only functions on V_m that achieve the maximum $(m-1)$th-order correlation immunity are $g(x_1, \ldots, x_m) = x_1 \oplus \cdots \oplus x_m$ and $g(x_1, \ldots, x_m) = x_1 \oplus \cdots \oplus x_m \oplus 1$, both of which are affine. For balanced functions, if $k \neq 0$ or $m - 1$, the relation becomes $k + d \leq m - 1$ [8].

Next we introduce a fundamental combinatorial structure, the *Hadamard matrix*. Properties of Hadamard matrices will be very useful in our constructions of correlation immune functions. A $(1, -1)$-matrix H of order m is called a Hadamard matrix if $HH^T = mI_m$, where H^T indicates the transpose of H and I_m is the identity matrix of order m. It is well known that the order m of an Hadamard matrix is 1, 2 or divisible by 4 [9, 6]. In this paper we will use a special kind of Hadamard matrices called *Sylvester-Hadamard matrices* or *Walsh-Hadamard matrices*. A Sylvester-Hadamard matrix (or Walsh-Hadamard matrix) of order 2^m, denoted by H_m, is generated by the following recursive relation

$$H_0 = 1, H_m = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \otimes H_{m-1}, m = 1, 2, \ldots$$

where $\otimes$ denotes the Kronecker product. Note that H_m can be written as $H_m = H_s \otimes H_t$ for any nonnegative integers s and t with $s + t = m$. Sylvester-Hadamard matrices are closely related to linear functions, as is shown in the following lemma.

Lemma 2 *Write* $H_m = \begin{bmatrix} \ell_0 \\ \ell_1 \\ \vdots \\ \ell_{2^m-1} \end{bmatrix}$ *where* ℓ_i *is a row of* H_m. *Then* ℓ_i *is the sequence of a linear function* $h_i = \langle \alpha_i, x \rangle$, *where* $x = (x_1, \ldots, x_m)$ *and* α_i *is a vector in* V_m *as defined in the first paragraph of this Section. Conversely the sequence of any linear function on* V_m *is a row of* H_m.

A proof for the first half of the lemma can be found in [7]. The second half is true by noting the fact that H_m has 2^m distinct rows and that there are exactly 2^m distinct linear functions on V_m. Thus the rows of $\pm H_m$ comprise all the *affine* sequences of length 2^m.

Next we introduce a notation which is used throughout the rest of the paper. Given any vector $\delta = (i_1, \ldots, i_s) \in V_s$, we define a function on V_s by

$$D_\delta(y) = (y_1 \oplus \bar{i_1}) \cdots (y_s \oplus \bar{i_s})$$

where $y = (y_1, \ldots, y_s)$ and $\bar{i} = 1 \oplus i$ indicates the binary complement of i. Note that since $D_\delta(y) = 1$ if and only if $y = \delta$, a function f on V_{s+t} can be expressed as

$$f(y, x) = \bigoplus_{\delta \in V_s} D_\delta(y) f(\delta, x)$$

where $x = (x_1, \ldots, x_t)$.

Lemma 3 *Let* $f(y,x) = \bigoplus_{\delta \in V_s} D_\delta(y) f_\delta(x)$ *and* $g(y,x) = \bigoplus_{\delta \in V_s} D_\delta(y) g_\delta(x)$ *where* $y = (y_1, \ldots, y_s)$, *and* $x = (x_1, \ldots, x_t)$. *Then* $f = g$ *if and only if* $f_\delta = g_\delta$ *for all* $\delta \in V_s$.

Proof. $f = g$ if and only if $f(\delta, x) = g(\delta, x)$ for all $\delta \in V_s$. Note that since $D_\delta(y) = 1$ if and only if $y = \delta$, we have $f(\delta, x) = f_\delta(x)$ and $g(\delta, x) = g_\delta(x)$ for all $\delta \in V_s$. □

The following lemma can be found in [7].

Lemma 4 *Let* $\xi_{i_1 \cdots i_p}$, $(i_1, \ldots, i_p) \in V_p$, *be the sequence of a function* $f_{i_1 \cdots i_p}(x_1, \ldots, x_q)$ *on* V_q. *Let* ξ *be the concatenation of* $\xi_{0\cdots00}$, $\xi_{0\cdots01}$, …, $\xi_{1\cdots11}$, *namely,* $\xi = (\xi_{0\cdots00}, \xi_{0\cdots01}, \ldots, \xi_{1\cdots11})$. *Then* ξ *is the sequence of a function on* V_{q+p} *given by*

$$f(y_1, \ldots, y_p, x_1, \ldots, x_q) = \bigoplus_{(i_1 \cdots i_p) \in V_p} D_{i_1 \cdots i_p}(y_1, \ldots, y_p) f_{i_1 \cdots i_p}(x_1, \ldots, x_q).$$

Let $\alpha = (a_1, a_2, \ldots, a_n) \in V_n$ and $\beta = (b_1, b_2, \ldots, b_m) \in V_m$. The *Kronecker product* of α and β, denoted by $\alpha \otimes \beta$, is defined as $\alpha \otimes \beta = (a_1\beta, a_2\beta, \ldots, a_m\beta)$. The following lemma will be used in the rest of the paper.

Lemma 5 *Let* ξ *be the sequence (or truth table) of a function* f *on* V_n *and* η *be the sequence (or truth table) of a function* g *on* V_m. *Then* $\xi \otimes \eta$ *is the sequence (or truth table) of the function* $\varphi(y, x) = f(y) \oplus g(x)$ *on* V_{n+m}.

Proof. For any fixed $y = \alpha \in V_n$, we have $\varphi(\alpha, x) = f(\alpha) \oplus g(x)$. □

The propagation characteristic is another nonlinearity measure for cryptographic functions. A function satisfies the propagation criterion of order k if complementing k or less input coordinates results in the output being complemented half the times over all input vectors. The formal definition for the propagation criterion follows.

Definition 2 *Let f be a function on V_n. We say that f satisfies*

1. *the* propagation criterion with respect to a non-zero vector α in V_n *if $f(x) \oplus f(x \oplus \alpha)$ is a balanced function.*

2. *the* propagation criterion of degree k *if it satisfies the propagation criterion with respect to all $\alpha \in V_n$ with $1 \leqq W(\alpha) \leqq k$.*

3 Previous Constructions

Siegenthaler presented a recursive construction in his pioneering work [8]. Let f_1 and f_2 be kth-order correlation immune functions on V_m. Then the concatenation of their sequences results in a new correlation immune function, namely,

$$f(u, x) = (u \oplus 1) f_1(x) \oplus u f_2(x) \tag{1}$$

is a kth-order correlation immune function on V_{m+1}, where u is a variable on $GF(2)$ and $x = (x_1, x_2, \ldots, x_m)$.

Camion et al [2] observed that in Siegenthaler's construction, if the Walsh transformations of f_1 and f_2 satisfy the condition

$$\hat{f}_1(\lambda) + \hat{f}_2(\lambda) = 0, \text{for all } \lambda \in V_m \text{ with } W(\lambda) = k,$$

then the order of the correlation immunity of f is improved to $k + 1$. In particular, they show the following two pairs of functions satisfy the condition:

1. $g(x)$ and $1 \oplus g(x)$;

2. $g(x)$ and $g(\bar{x})$, where $\bar{x} = (1 \oplus x_1, 1 \oplus x_2, \ldots, 1 \oplus x_m)$;

where g is a kth-order correlation immune function on V_m. Note that $1 \oplus g(x)$ complements the output, while $g(\bar{x})$ complements the input. Therefore, both

$$f(x) = (u \oplus 1) g(x) \oplus u(1 \oplus g(x)) = u \oplus g(x) \tag{2}$$

and

$$f(x) = (u \oplus 1) g(x) \oplus u g(\bar{x}) = g(x) \oplus u(g(x) \oplus g(\bar{x})) \tag{3}$$

are $(k + 1)$th-order correlation immune functions on V_{m+1}.

In the same paper, Camion et al also discovered a method for direct construction of correlation immune functions. Let m and n be positive integers with $m > n$. Let

r and p_j, $j = 1, 2, \ldots, n$ be arbitrary functions on V_{m-n}. Also let $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_{m-n})$. Set

$$f(y, x) = \bigoplus_{j=1}^{n} x_j p_j(y) \oplus r(y). \tag{4}$$

Then the function f defined in (4) is a balanced kth-order correlation immune function on V_m, where k is an integer satisfying $k \geqq \min\{W(P(y)) | y \in V_{m-n}\} - 1$, and $P(y) = (p_1(y), p_2(y), \ldots, p_n(y))$.

4 A New Construction

Let m and n be positive integers with $m > n$. Suppose that $\Phi_{m,n} = \{\varphi_{0\cdots0}, \varphi_{0\cdots1}, \ldots, \varphi_{1\cdots1}\}$ is a set containing 2^{m-n} linear functions on V_n, each is indexed by a vector in V_{m-n}. $\Phi_{m,n}$ can be a multi-set and hence a linear function is allowed to appear more than once in $\Phi_{m,n}$. Let $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_{m-n})$ and r be an arbitrary function on V_{m-n}. Set

$$g(y, x) = \bigoplus_{\delta \in V_{m-n}} D_\delta(y) \varphi_\delta(x) \oplus r(y) \tag{5}$$

The following corollary is a consequence of Theorem 1 and Corollary 2 to be stated below, though it can be proved directly.

Corollary 1 *The function g defined in (5) is a balanced kth-order correlation immune function on V_m, where k is an integer satisfying $k \geqq \min\{W(\gamma_\delta) | \delta \in V_{m-n}\} - 1$, $\varphi_\delta(x) = \langle \gamma_\delta, x \rangle \in \Phi_{m,n}$ and $\gamma_\delta \in V_n$.*

Theorem 1 *The constructions (4) and (5) express the same set of functions.*

Proof. Let S_1 be the set of functions generated by (4) and S_2 the set of functions generated by (5).

First we prove that $S_1 \subseteq S_2$ by showing that a function obtained by (4) can always be represented in the form of (5). Let

$$f(y, x) = \bigoplus_{j=1}^{n} x_j p_j(y) \oplus r(y)$$

be a function in S_1. For any $\delta \in V_{m-n}$ we have

$$f(\delta, x) = \bigoplus_{j=1}^{n} x_j p_j(\delta) \oplus r(\delta).$$

Since $p_j(\delta) \in GF(2)$, $j = 1, \ldots, n$, $\bigoplus_{j=1}^{n} x_j p_j(\delta)$ is a linear function on V_n. Now let

$$\varphi_\delta(x) = \bigoplus_{j=1}^{n} x_j p_j(\delta),$$

and let

$$g(y,x) = \bigoplus_{\delta \in V_{m-n}} D_\delta(y)\varphi_\delta(x) \oplus r(y).$$

Note that $D_\delta(y) = 1$ if and only if $y = \delta$. Thus we have

$$g(\delta, x) = \varphi_\delta(x) \oplus r(\delta) = f(\delta, x).$$

Since δ is arbitrary, by Lemma 3 we have

$$f(y,x) = g(y,x).$$

Consequently, $f(y,x)$ can be represented in the form of (5). This means that $S_1 \subseteq S_2$.

Next we show that a function obtained by (5) can be represented in the form of (4). This will prove that $S_2 \subseteq S_1$. Let

$$g(y,x) = \bigoplus_{\delta \in V_{m-n}} D_\delta(y)\varphi_\delta(x) \oplus r(y)$$

be a function in S_2. Let δ be an arbitrary vector in V_{m-n}, and let

$$\varphi_\delta(x) = a_{\delta,1}x_1 \oplus \cdots \oplus a_{\delta,n}x_n \tag{6}$$

Now let p_j, $j = 1, 2, \ldots, n$, be a function on V_{m-n} such that

$$p_j(\delta) = a_{\delta,j}$$

for all $\delta \in V_{m-n}$. Also let $P = (p_1, \ldots, p_n)$ be a mapping from V_{m-n} to V_n such that

$$P(\delta) = (p_1(\delta), \ldots, p_n(\delta)) \tag{7}$$

for all $\delta \in V_{m-n}$. Now we define a function on V_m in the following way

$$f(y,x) = \bigoplus_{j=1}^{n} x_j p_j(y) \oplus r(y).$$

Again since $D_\delta(y) = 1$ if and only if $y = \delta$, we have

$$g(\delta, x) = \varphi_\delta(x) \oplus r(\delta).$$

By (6) and (7) we have

$$f(\delta,x) = \bigoplus_{j=1}^{n} x_j p_j(\delta) \oplus r(\delta) = \bigoplus_{j=1}^{n} x_j a_{\delta,j} \oplus r(\delta) = \varphi_\delta(x) \oplus r(\delta) = g(\delta,x).$$

Since δ is arbitrary, by Lemma 3 we have

$$g(y,x) = f(y,x).$$

This implies that $g(y,x)$ can be presented in the form of (4) and thus $S_2 \subseteq S_1$. This completes the proof that $S_1 = S_2$. □

Corollary 2 *In the proof of Theorem 1*

$$\min\{W(P(y))|y \in V_{m-n}\} - 1 = \min\{W(\gamma_\delta)|\delta \in V_{m-n}\} - 1.$$

where $\varphi_\delta(x) = \langle \gamma_\delta, x\rangle = a_{\delta,1}x_1 \oplus \cdots \oplus a_{\delta,n}x_n$ *and* $\gamma_\delta = (a_{\delta,1}, \ldots, a_{\delta,n})$ *are the same as in the proof of Theorem 1.*

Proof. From (7) we have $P(\delta) = (a_{\delta,1}, \ldots, a_{\delta,n})$, and from (6) we have $\varphi_\delta(x) = a_{\delta,1}x_1 \oplus \cdots \oplus a_{\delta,n}x_n, = \langle \gamma_\delta, x\rangle$. Thus we have $P(\delta) = \gamma_\delta$ and hence $\min\{W(P(y))|y \in V_{m-n}\} - 1 = \min\{W(\gamma_\delta)|\delta \in V_{m-n}\} - 1$. □

5 Applying the New Construction

For integers k and n with $0 \leqq k < n$, let $\Omega_{k,n}$ denote the set of linear functions on V_n that have $k+1$ or more non-zero coefficients, namely

$$\Omega_{k,n} = \{\varphi|\varphi(x) = \langle \beta, x\rangle, \beta \in V_n, W(\beta) \geqq k+1\} \tag{8}$$

where $x = (x_1, \ldots, x_n)$. This set of functions will be used in our constructions of correlation immune functions.

5.1 Balanced Functions with Given Immunity

Given two integers m and k with $m \geqq 3$ and $1 \leqq k < m-1$, balanced kth-order correlation immune functions on V_m can be constructed in the following way.

1. Fix an integer n such that $k < n < m$.
2. Create a set $\Phi_{m,n}$ by selecting linear functions strictly from $\Omega_{k,n}$. Note that the size of $\Phi_{m,n}$ is 2^{m-n}, and repetition is permitted in the selection.
3. Construct a function by using the method (5).

By Corollary 1, we have

Theorem 2 *A function constructed according to the above three steps is a balanced* k*th-order correlation immune function on* V_m*.*

5.2 Algebraic Degrees

Let k and m be integers with $k \geqq 1$ and $m \geqq k+2$. As mentioned in Section 2, the algebraic degree of a balanced kth-order immune correlation functions on V_m is at most $m-k-1$. We are interested in constructing balanced kth-order correlation immune functions having the maximum algebraic degree $m-k-1$.

In order to discuss their algebraic degrees, we construct functions in the following three steps.

1. Fix an integer n such that $m > n \geqq k+2$.

2. Choose a multi-set $\Phi_{m,n} = \{\varphi_\delta : V_n \to GF(2) | \delta \in V_{m-n}\}$ of linear functions in such a way that it satisfies the following three conditions:

   (C1) If $\varphi \in \Phi_{m,n}$ then $\varphi \in \Omega_{k,n}$, where $\Omega_{k,n}$ is defined in (8),

   (C2) $\Phi_{m,n}$ contains at least two distinct functions,

   (C3) there is a variable x_j that appears in an odd number of functions in $\Phi_{m,n}$. Note that the repetition of functions is counted by the number of appearance.

3. Employ the set $\Phi_{m,n}$ in the construction (5).

Since $\Phi_{m,n}$ is a multi-set, the condition (C1) can be satisfied. On the other hand, since $n \geqq k+2$ and $\Omega_{k,n}$ contains more than two functions, the condition (C2) can also be readily satisfied.

Once the conditions (C1) and (C2) are satisfied, we check $\Phi_{m,n}$ to see if it satisfies the condition (C3). If not, we modify $\Phi_{m,n}$ in the following way. Since $\Phi_{m,n}$ satisfies the condition (C2), there are two distinct functions $\varphi_{\delta_1}(x), \varphi_{\delta_2}(x) \in \Phi_{m,n}$. Thus there exists some x_j that appears in $\varphi_{\delta_1}(x)$ but not in $\varphi_{\delta_2}(x)$. Now we replace $\varphi_{\delta_2}(x)$ by $\varphi_{\delta_1}(x)$. In this way we can modify the function set $\Phi_{m,n}$ so that it satisfies the condition (C3). When the condition (C3) is satisfied, there is a term $y_1 \cdots y_{m-n} x_j$ that appears an odd number of times in a function g constructed according to the above three steps. This term survives in the final algebraic normal form representation of g. In other words, the algebraic degree of g is $m-n+1$.

From Theorem 2 and the above discussions, we know that g is a balanced kth-order correlation immune function of algebraic degree $m-n+1$. Thus we have proved

Theorem 3 *Let k, n and m be integers with $k \geqq 1$ and $m > n \geqq k+2$. Then a function constructed according to the above three steps is a balanced kth-order correlation immune function on V_m of algebraic degree $m-n+1$. When n is chosen as $n = k+2$, the function achieve the maximum algebraic degree $m-k-1$.*

5.3 Nonlinearity

Given two functions f and g on V_m, the *Hamming distance* between f and g is defined as $d(f,g) = W(f(x) \oplus g(x))$. The *nonlinearity* of g is defined as $N_f = \min_{i=0,1,\ldots,2^{m+1}-1} d(f,\varphi)$ where φ_0, φ_1, ..., $\varphi_{2^{m+1}-1}$ comprise all the affine functions on V_m. It has been proved that $N_f \leqq 2^{m-1} - 2^{\frac{m}{2}-1}$ for any function f on V_m [7]. Nonlinearity is an crucial criterion for cryptographic functions and it measures the ability of a cryptographic system using the functions to resist being expressed as a set of linear equations. If the system could be expressed as linear equations, it would be easily breakable by various attacks.

Let f_1 and f_2 be functions on V_m, ξ_1 and ξ_2 be the sequences of f_1 and f_2 respectively. Then $\langle \xi_f, \xi_g \rangle = \sum_{f(x)=g(x)} 1 - \sum_{f(x)\neq g(x)} 1 = 2^m - 2\sum_{f(x)\neq g(x)} 1 = 2^m - 2d(f,g)$. This proves the following result which is very useful in the study of the nonlinearity of functions.

Lemma 6 *Let f and g be functions on V_m whose sequences are ξ_f and ξ_g respectively. Then the distance between f and g can be calculated by $d(f,g) = 2^{m-1} - \frac{1}{2}\langle \xi_f, \xi_g \rangle$.*

Now we calculate the nonlinearity of correlation immune functions constructed by (5).

Theorem 4 *Let m and n be integers with $m > n > 2$, and let g be a function constructed by (5). Denote by t_δ the number of times a linear function φ_δ appears in $\Phi_{m,n}$, and let $t = \max\{t_\delta | \delta \in V_{m-n}\}$. Then the nonlinearity of g satisfies $N_g \geqq 2^{m-1} - t2^{n-1}$.*

Proof. For convenience a vector $\delta \in V_{m-n}$ will be denoted by its corresponding integer between 0 and $2^{m-n}-1$. In this way, a linear function $\varphi_\delta \in \Phi_{m,n}$ indexed by δ is rewritten as φ_j and t_δ is rewritten as t_j, where t_δ is the number of times φ_δ appears in $\Phi_{m,n}$ and j is the integer representation of δ. We first consider the case when $r(y) = 0$ in the construction (5), namely

$$g(y,x) = D_{0\cdots 0}(y)\varphi_0(x) \oplus \cdots \oplus D_{1\cdots 1}(y)\varphi_{2^{m-n}-1}(x) \tag{9}$$

where $\varphi_j \in \Omega_{k,n}$, $y = (y_1, \ldots, y_{m-n})$, $x = (x_1, \ldots, x_n)$, and $D_{j_1\cdots j_{m-n}}$ is defined in Section 2.

Let h be any affine function on V_m. By Lemma 2, the sequence of h, denoted by L, is a row of $\pm H_m$. Since $H_m = H_{m-n} \otimes H_n$, L can be expressed as $L = \pm \ell' \otimes \ell''$, the Kronecker product of ℓ' and ℓ'', where ℓ' is a row of H_{m-n} while ℓ'' is a row of H_n. Write ℓ' as $\ell' = (c_0, c_1, \ldots, c_{2^{m-n}-1})$. Then L can be rewritten as $L = (c_0\ell'', c_1\ell'', \ldots, c_{2^{m-n}-1}\ell'')$. Note that by Lemma 2, ℓ'' is the sequence of a linear function. We denote the linear function by φ''.

Now let ζ_j be the sequence of φ_j, $j = 0, 1, \ldots, 2^{m-n}-1$. By Lemma 4, $\eta = (\zeta_0, \zeta_1, \ldots, \zeta_{2^{m-n}-1})$ is the sequence of g defined in (9). On the other hand, since the rows of an Hadamard matrix are mutually orthogonal, we have the following result:

$$\langle \zeta_j, \ell'' \rangle = \begin{cases} 2^n, & \text{if } \varphi_j = \varphi'' \\ 0, & \text{otherwise.} \end{cases}$$

Now we discuss $\langle \eta, L \rangle$ in the following two cases:

Case 1: there exists a j such that $\varphi_j = \varphi''$; since φ_j appears t_j times in $\Phi_{m,n}$, the total number of times when $\varphi_j = \varphi''$ is also t_j. Thus $|\langle \eta, L \rangle| \leqq t_j 2^n$.

Case 2: there exists no j such that $\varphi_j = \varphi''$; in this case we have $|\langle \eta, L \rangle| = 0$.

Summarizing Cases 1 and 2, we have $|\langle \eta, L \rangle| \leqq t2^n$. By Lemma 6, $d(g,h) \geqq 2^{m-1} - t2^{n-1}$. Since h is arbitrary, we have $N_g \geqq 2^{m-1} - t2^{n-1}$.

Now consider the more general case when $r(y) \neq 0$ in the construction (5). Since r is a function of y but not x, the sequence of g takes the form of $\eta = (e_0\zeta_0, e_1\zeta_1, \ldots, e_{2^{m-n}-1}\zeta_{2^{m-n}-1})$, where $e_i = (-1)^{r(\alpha_i)}$ and α_i is a vector in V_{m-n} whose integer representation is i. By a similar discussion to the case when $r(y) = 0$, we have $|\langle \eta, L \rangle| \leqq t2^n$ for any affine sequence L, and hence $N_g \geqq 2^{m-1} - t2^{n-1}$. □

5.4 Propagation Characteristics

This section discusses the propagation characteristics of functions obtained by (5). For convenience, the construction method is repeated here:

$$g(y,x) = \bigoplus_{\delta \in V_{m-n}} D_\delta(y)\varphi_\delta(x) \oplus r(y)$$

In the following discussion, we assume that all linear functions φ_δ in the construction are distinct.

It is easy to prove that

$$D_\delta(y \oplus \beta) = D_{\delta \oplus \beta}(y).$$

Let $z = (y, x)$. Also let $\beta \in V_{m-n}$, $\alpha \in V_n$ and $\gamma = (\beta, \alpha)$. Then

$$\begin{aligned} g(z \oplus \gamma) &= \bigoplus_{\delta \in V_{m-n}} D_\delta(y \oplus \beta)\varphi_\delta(x \oplus \alpha) \oplus r(y \oplus \beta) \\ &= \bigoplus_{\delta \in V_{m-n}} (y) D_{\delta \oplus \beta}(y)\varphi_\delta(x \oplus \alpha) \oplus r(y \oplus \beta) \\ &= \bigoplus_{\delta \oplus \beta \in V_{m-n}} D_{\delta \oplus \beta}(y)\varphi_\delta(x \oplus \alpha) \oplus r(y \oplus \beta) \end{aligned}$$

Set $\sigma = \delta \oplus \beta$, we have

$$g(z \oplus \gamma) = \bigoplus_{\sigma \in V_{m-n}} D_\sigma(y)\varphi_{\sigma \oplus \beta}(x \oplus \alpha) \oplus r(y \oplus \beta)$$

and hence

$$g(z) \oplus g(z \oplus \gamma) = \bigoplus_{\sigma \in V_{m-n}} D_\sigma(y)(\varphi_\sigma(x) \oplus \varphi_{\sigma \oplus \beta}(x \oplus \alpha)) \oplus r(y) \oplus r(y \oplus \beta).$$

Note that for any fixed $y = \sigma$

$$(g(z) \oplus g(z \oplus \gamma))|_{y=\delta} = \varphi_\sigma(x) \oplus \varphi_{\sigma \oplus \beta}(x \oplus \alpha) \oplus r(\sigma) \oplus r(\sigma \oplus \beta).$$

Consider the case when $\beta \neq (0, \ldots, 0)$. By assumption $\varphi_\sigma(x)$ and $\varphi_{\sigma \oplus \beta}(x)$ are distinct linear functions. Hence $\varphi_\sigma(x) \oplus \varphi_{\sigma \oplus \beta}(x \oplus \alpha)) = \varphi_\sigma(x) \oplus \varphi_{\sigma \oplus \beta}(x) \oplus \varphi_{\delta \oplus \beta}(\alpha)$ is a non-constant affine function which is balanced. This shows that $g(z) \oplus g(z \oplus \gamma)$ is balanced for any $\gamma = (\beta, \alpha)$ with $\beta \neq (0, \ldots, 0)$. Thus we have proved

Theorem 5 *In the construction (5), if all φ_δ are distinct linear functions on V_n, then g satisfies the propagation criterion with respect to all γ with $\gamma = (\beta,\ \alpha)$, $\beta \in V_{m-n}$, $\alpha \in V_n$ and $\beta \neq 0$.*

Note that there are $2^{m-n} - 1$ choices for $\beta \neq 0$ and 2^n choices for all $\alpha \in V_n$. Therefore the total number of vectors with respect to which the function g satisfies the propagation criterion is at least $(2^{m-n} - 1)2^n = 2^m - 2^n$.

5.5 Examples

Theorem 3 gives us a general method to construct balanced correlation immune functions having any given immunity. The construction method allows us to easily calculate the algebraic degree and the nonlinearity of the functions, which is very desirable in designing cryptographic systems. Two concrete examples follow.

Let $n = 4$ and $k = 2$. Then

$$\begin{aligned}\Omega_{2,4} &= \{\varphi | \varphi(x) = \langle \beta, x \rangle, \beta \in V_4, W(\beta) \geqq 3\} \\ &= \{x_1 \oplus x_2 \oplus x_3,\ x_1 \oplus x_2 \oplus x_4,\ x_1 \oplus x_3 \oplus x_4,\ x_2 \oplus x_3 \oplus x_4,\ x_1 \oplus x_2 \oplus x_3 \oplus x_4\}.\end{aligned}$$

where $x = (x_1, x_2, x_3, x_4)$.

Example 1 We construct a balanced 2nd-order immune function f on V_7, which achieves the maximum algebraic degree of 4. We also calculate the nonlinearity of the function.

Set

$$\begin{aligned}\varphi_1(x) &= x_1 \oplus x_2 \oplus x_3, & \varphi_5(x) &= \varphi_1(x) \\ \varphi_2(x) &= x_1 \oplus x_2 \oplus x_4, & \varphi_6(x) &= \varphi_2(x) \\ \varphi_3(x) &= x_1 \oplus x_3 \oplus x_4, & \varphi_7(x) &= \varphi_3(x) \\ \varphi_4(x) &= x_2 \oplus x_3 \oplus x_4, & \varphi_8(x) &= \varphi_3(x)\end{aligned}$$

and

$$\Phi_{7,4} = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7, \varphi_8\}.$$

$\Phi_{7,4}$ is a multi-set whose elements are all taken from $\Omega_{2,4}$. In addition, it contains four different functions, and x_1 appears in seven functions. Thus the three conditions (C1), (C2) and (C3) are all satisfied.

To complete the construction, let

$$\begin{aligned}f(y,x) &= D_{000}(y)\varphi_1(x) \oplus D_{001}(y)\varphi_2(x) \oplus D_{010}(y)\varphi_3(x) \oplus D_{011}(y)\varphi_4(x) \oplus \\ &\quad D_{100}(y)\varphi_5(x) \oplus D_{101}(y)\varphi_6(x) \oplus D_{110}(y)\varphi_7(x) \oplus D_{111}(y)\varphi_8(x) \\ &= (1 \oplus y_2y_3 \oplus y_1y_2y_3)x_1 \oplus (1 \oplus y_2 \oplus y_2y_3 \oplus y_1y_2y_3)x_2 \oplus \\ &\quad (1 \oplus y_3 \oplus y_2y_3)x_3 \oplus (y_2 \oplus y_3 \oplus y_2y_3)x_4\end{aligned}$$

where $y = (y_1, y_2, y_3)$ and $x = (x_1, x_2, x_3, x_4)$.

By Theorem 3, f is a balanced 2nd-order correlation immune function on V_7 of algebraic degree 4. To calculate the nonlinearity of the function, note that $\varphi_3 = \varphi_7 = \varphi_8$ and hence $t = \max\{t_j | j = 1, \ldots, 8\} = 3$. By Theorem 4, we have $N_f \geqq 2^{7-1} - 3 \cdot 2^{4-1} = 40$. Note that the upper bound of the nonlinearity of balanced functions on V_7 is 56 (see Corollary 17 of [7].

Example 2 In this example, we construct a balanced 2nd-order immune function g on V_6. Let

$$\begin{aligned}\varphi_1(x) &= x_1 \oplus x_2 \oplus x_3, \\ \varphi_2(x) &= x_1 \oplus x_2 \oplus x_4, \\ \varphi_3(x) &= x_1 \oplus x_3 \oplus x_4, \\ \varphi_4(x) &= x_1 \oplus x_2 \oplus x_3 \oplus x_4,\end{aligned}$$

and

$$\Phi_{6,4} = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4\}.$$

Obviously $\Phi_{6,4}$ satisfies the three conditions (C1), (C2) and (C3).

Let

$$\begin{aligned} g(y,x) &= D_{00}(y)\varphi_1(x) \oplus D_{01}(y)\varphi_2(x) \oplus D_{10}(y)\varphi_3(x) \oplus D_{11}(y)\varphi_4(x) \\ &= x_1 \oplus (1 \oplus y_1 \oplus y_1y_2)x_2 \oplus \\ &\quad (1 \oplus y_2 \oplus y_1y_2)x_3 \oplus (y_1 \oplus y_2 \oplus y_1y_2)x_4 \end{aligned}$$

where $y = (y_1, y_2)$ and $x = (x_1, x_2, x_3, x_4)$.

g is a balanced 2nd-order correlation immune function on V_6. It satisfies the propagation criterion with respect to all $\alpha = (a_1, a_2, a_3, a_4, a_5, a_6) \in V_6$ with $a_1 \neq 0$ or $a_2 \neq 0$. The algebraic degree of g is 3 and the nonlinearity of g is $N_g \geqq 2^{6-1} - 2^{4-1} = 24$. For comparison, note that the upper bound for the nonlinearity of balanced functions on V_6 is 26 (see [7]).

6 Combination of Correlation Immune Functions

The construction (5) described in Section 4 presents a method for directly constructing correlation immune functions of any order. In this section we discuss three methods for constructing correlation immune functions on a higher dimensional space from existing such functions on a lower dimensional space.

6.1 An Extension of the New Construction

The construction (5) can be extended. Let m, n, k and s be positive integers, where $m > n > k$, and let $w = (y, x, z)$, $y = (y_1, \ldots, y_{m-n})$, $x = (x_1, \ldots, x_n)$ and $z = (z_1, \ldots, z_s)$. Also let $\Phi_{m,n} = \{\varphi_0, \ldots, \varphi_{2^{m-n}-1}\}$ be a set of linear functions on V_n, each of which is selected from $\Omega_{k,n}$. Repetition is permitted in selecting the linear functions. Set

$$g_1(y,x) = D_{0\cdots0}(y)\varphi_0(x) \oplus \cdots \oplus D_{1\cdots1}(y)\varphi_{2^{m-n}-1}(x) \oplus r_1(y) \tag{10}$$

where r_1 is an arbitrary function on V_{m-n}. By Corollary 1, g_1 is a balanced kth-order correlation immune functions on V_m.

Now let $\{f_0, \ldots, f_{2^{m-n}-1}\}$ be a set of pth-order correlation immune functions on V_s. Functions in the set need not be mutually distinct. Set

$$g_2(y,z) = D_{0\cdots0}(y)f_0(z) \oplus \cdots \oplus D_{1\cdots1}(y)f_{2^{m-n}-1}(z) \oplus r_2(y) \tag{11}$$

where r_2 is an arbitrary function on V_{m-n}. We further set

$$g(y,x,z) = g_1(y,x) \oplus g_2(y,z) \tag{12}$$

Theorem 6 *The function $g(y,x,z) = g_1(y,x) \oplus g_2(y,z)$ is a balanced $(k+p+1)$th-order correlation immune function on V_{m+s}. The nonlinearity of g satisfies*

$$N_g \geqq 2^{m-1} - t \cdot 2^n (2^{s-1} - N)$$

where $t = \max\{t_j | j = 0, 1, \ldots, 2^{m-n}-1\}$, t_j denotes the number of times that φ_j appears in $\Phi_{m,n}$, and $N = \min\{N_{f_j} | j = 0, 1, \ldots, 2^{m-n}-1\}$.

Proof. We first consider the case when $r(y) = r_1(y) \oplus r_2(y) = 0$. Note that

$$g(y,x,z) = D_{0\cdots0}(y)(\varphi_0(x) \oplus f_0(z)) \oplus \cdots \oplus D_{1\cdots1}(y)(\varphi_{2^{m-n}-1}(x) \oplus f_{2^{m-n}-1}(z)).$$

Since each φ_j is balanced, each $\varphi_j(x) \oplus f_j(z)$ is also balanced (see Lemma 20 of [7]). Hence $g(y,x,z)$ is balanced.

Now we show that g is a $(k+p+1)$th-order correlation immune function. Let ζ_j and ξ_j be the sequences of φ_j and f_j respectively, $j = 0, 1, \ldots, 2^{m-n}-1$. By Lemma 5 $\zeta_j \otimes \xi_j$ is the sequence of $\varphi_j(x) \oplus f_j(z)$, and $\eta = (\zeta_0 \otimes \xi_0, \ldots, \zeta_{2^{m-n}-1} \otimes \xi_{2^{m-n}-1})$ is the sequence of $g(y,x,z)$ (see Lemma 4).

Let h be a linear function on V_{m+s}. By Lemma 2, the sequence of h, denoted by L, is a row of H_{m+s}. Since $H_{m+s} = H_{m-n} \otimes H_n \otimes H_s$, L can be expressed as $L = \ell_1 \otimes \ell_2 \otimes \ell_3$, where ℓ_1 is a row of H_{m-n}, ℓ_2 is a row of H_n, and ℓ_3 is a row of H_s. Write $\ell_1 = (c_0, c_1, \ldots, c_{2^{m-n}-1})$. Then L can be rewritten as $L = (c_0\ell_2 \otimes \ell_3, \ldots, c_{2^{m-n}-1}\ell_2 \otimes \ell_3)$. Let η be the sequence of g. Then

$$\begin{aligned} \langle \eta, L \rangle &= c_0 \langle \zeta_0 \otimes \xi_0, \ell_2 \otimes \ell_3 \rangle + \cdots + c_{2^{m-n}-1} \langle \zeta_{2^{m-n}-1} \otimes \xi_{2^{m-n}-1}, \ell_2 \otimes \ell_3 \rangle \\ &= c_0 \langle \zeta_0, \ell_2 \rangle \langle \xi_0, \ell_3 \rangle + \cdots + c_{2^{m-n}-1} \langle \zeta_{2^{m-n}-1}, \ell_2 \rangle \langle \xi_{2^{m-n}-1}, \ell_3 \rangle. \end{aligned}$$

Write $h(w) = \langle \gamma, w \rangle = \langle \beta, y \rangle \oplus \langle \alpha, x \rangle \oplus \langle \sigma, z \rangle$, where $\gamma = (\beta, \alpha, \sigma)$, $\beta \in V_{m-n}$, $\alpha \in V_n$ and $\sigma \in V_s$. By the definition of the sequence of a function, ℓ_1, ℓ_2 and ℓ_3 are the sequences of $\langle \beta, y \rangle$, $\langle \alpha, x \rangle$ and $\langle \sigma, z \rangle$ respectively.

Suppose that $W(\gamma) \leqq k+p+1$. Since $W(\gamma) = W(\beta) + W(\alpha) + W(\sigma)$, we have $W(\alpha) + W(\sigma) \leqq k+p+1$, which implies that either $W(\alpha) \leqq k$ or $W(\sigma) \leqq p$. Recall that $\varphi_j \in \Omega_{k,n}$. If $W(\alpha) \leqq k$, ζ_j and ℓ_2 must be orthogonal, and hence $\langle \zeta_j, \ell_2 \rangle = 0$. Otherwise if $W(\sigma) \leqq p$, $\langle \xi_j, \ell_3 \rangle = 0$, since each f_j is a pth-order correlation immune function. Thus $\langle \eta, L \rangle = 0$. By Lemma 1, $g(y,x,z)$ is a $(k+p+1)th$-order correlation immune function on V_{m+s}.

To obtain the nonlinearity of the function g, we assume that in the above discussion h is an arbitrary affine function on V_{m+s}. Then L, the sequence of h, can be expressed as $L = \pm\ell_1 \otimes \ell_2 \otimes \ell_3$, and hence

$$\langle \eta, L \rangle = \pm(c_0 \langle \zeta_0, \ell_2 \rangle \langle \xi_0, \ell_3 \rangle + \cdots + c_{2^{m-n}-1} \langle \zeta_{2^{m-n}-1}, \ell_2 \rangle \langle \xi_{2^{m-n}-1}, \ell_3 \rangle).$$

By Lemma 5

$$\langle \xi_j, \ell_3 \rangle \leqq 2^s - 2N_{f_j} \leqq 2^s - 2N.$$

On the other hand, since the rows of an Hadamard matrix are mutually orthogonal, we have the following result:

$$\langle \zeta_j, \ell_2 \rangle = \begin{cases} 2^n & \text{if } \zeta_j = \ell_2, \\ 0 & \text{otherwise.} \end{cases}$$

When there is a j such that $\zeta_j = \ell_2$, we have $|\langle \eta, L \rangle| \leqq t \cdot 2^n(2^s - 2N)$. Otherwise if there is no j such that $\zeta_j = \ell_2$, $|\langle \eta, L \rangle| = 0$. In summary, we have $|\langle \eta, L \rangle| \leqq t \cdot 2^n(2^s - 2N)$. By Lemma 5, $d(g,h) \geqq 2^{m-1} - t \cdot 2^n(2^{s-1} - N)$. Since h is arbitrary, $N_g \geqq 2^{m-1} - t \cdot 2^n(2^{s-1} - N)$.

By a similar discussion as in the last part of the proof of Theorem 4, the theorem is true for the more general case when $r(y) = r_1(y) \oplus r_2(y) \neq 0$. □

The construction (12) can be considered as an extension of the construction (5), in the sense that if $s = 0$ and each function f_j is defined as a constant, the former is reduced to the latter.

6.2 Direct Sum of Two Correlation Immune Functions

Lemma 7 *Let f_1 be a k_1th-order correlation immune function on V_{n_1}, f_2 be a k_2th-order correlation immune function on V_{n_2}. Then $g(x,y) = f_1(x) \oplus f_2(y)$ is a $(k_1 + k_2 + 1)$th-order correlation immune function on $V_{n_1+n_2}$, where $x = (x_1, x_2, \ldots, x_{n_1})$ and $y = (y_1, y_2, \ldots, y_{n_2})$.*

Proof. Let ξ_1 and ξ_2 be the sequences of f_1 and f_2 respectively. Then by Lemma 5, $\eta = \xi_1 \otimes \xi_2$ is the sequence of g.

Let φ be a linear function on $V_{n_1+n_2}$. Then φ can be written as $\varphi = \langle \gamma, z \rangle = \langle \alpha, x \rangle \oplus \langle \beta, y \rangle$, where $z = (x,y), \gamma = (\alpha, \beta) \in V_{n_1+n_2}$, $\alpha \in V_{n_1}$ and $\beta \in V_{n_2}$. Now let L be the sequence of φ. By Lemma 2, L is a row of $H_{n_1+n_2}$. Since $H_{n_1+n_2} = H_{n_1} \otimes H_{n_2}$, L can be expressed as $L = \ell_1 \otimes \ell_2$, where ℓ_1 is a row of H_{n_1} and ℓ_2 is a row of H_{n_2}.

Now we show that ℓ_1 matches the sequence of $\langle \alpha, x \rangle$, and ℓ_2 matches the sequence of $\langle \beta, y \rangle$. Assume that ℓ_1' is the sequence of $\langle \alpha, x \rangle$, and ℓ_2' is the sequence of $\langle \beta, y \rangle$. By Lemma 5, $\ell_1' \otimes \ell_2'$ is the sequence of φ. Thus $L = \ell_1 \otimes \ell_2 = \ell_1' \otimes \ell_2'$. By Lemma 2, ℓ_1' is a row of H_{n_1} and ℓ_2' is a row of H_{n_2}. This means that $\ell_1 = \ell_1'$ and $\ell_2 = \ell_2'$. Put it in another way, ℓ_1 is the sequence of $\langle \alpha, x \rangle$, and ℓ_2 is the sequence of $\langle \beta, y \rangle$.

Now consider γ with $W(\gamma) \leqq k_1 + k_2 + 1$. In this case we have either $W(\alpha) \leqq k_1$ or $W(\beta) \leqq k_2$. Thus

$$\langle \eta, L \rangle = \langle \xi_1 \otimes \xi_2, \ell_1 \otimes \ell_2 \rangle = \langle \xi_1, \ell_1 \rangle \langle \xi_2, \ell_2 \rangle = 0.$$

By Lemma 1, g is indeed a $(k_1 + k_2 + 1)$th-order correlation immune function on $V_{n_1+n_2}$. □

Lemma 8 *Let f_1 be a function on V_{n_1} and f_2 be a function on V_{n_2}. Suppose that their nonlinearities are $N_{f_1} = d_1$ and $N_{f_2} = d_2$ respectively. Then the nonlinearity of $g(x,y) = f_1(x) \oplus f_2(y)$ satisfies $N_g \geqq d_1 2^{n_2} + d_2 2^{n_1} - 2d_1 d_2$.*

Proof. Let ξ_1, ξ_2, η, L, ℓ_1, ℓ_2, φ be the same as in the proof of Lemma 7. Let $\varphi_1 = \langle \alpha, x \rangle$ and $\varphi_2 = \langle \beta, y \rangle$.

By Lemma 6, we have

$$d_1 = N_{f_1} \leqq d(f_1, \varphi_1) = 2^{n_1 - 1} - \frac{1}{2} \langle \xi_1, \ell_1 \rangle.$$

Thus

$$\langle \xi_1, \ell_1 \rangle \leqq 2^{n_1} - 2d_1. \tag{13}$$

Similarly

$$\langle \xi_2, \ell_2 \rangle \leqq 2^{n_2} - 2d_2. \tag{14}$$

Note that the right sides of (13) and (14) are both positive. Thus

$$\langle \eta, L \rangle = \langle \xi_1 \otimes \xi_2, \ell_1 \otimes \ell_2 \rangle = \langle \xi_1, \ell_1 \rangle \langle \xi_2, \ell_2 \rangle \leqq (2^{n_1} - 2d_1)(2^{n_2} - 2d_2). \tag{15}$$

Again by Lemma 6,

$$d(g, \varphi) = 2^{n_1+n_2-1} - \frac{1}{2}\langle \eta, L \rangle \geqq d_1 2^{n_2} + d_2 2^{n_1} - 2d_1 d_2.$$

It is easy to see that the right side of (15) is also positive. Thus if L is an *affine* sequence (i.e. φ is an affine function) (15) still holds. Since φ is an arbitrary affine function we have

$$N_g \geqq d_1 2^{n_2} + d_2 2^{n_1} - 2d_1 d_2.$$

Therefore the lemma is true. □

Combining Lemmas 7 and 8 and using Lemma 20 of [7] we have

Theorem 7 *Let f_1 be a k_1th-order correlation immune function on V_{n_1} and f_2 be a k_2th-order correlation immune function on V_{n_2}. Also suppose that $N_{f_1} = d_1$ and $N_{f_2} = d_2$. Then $g(x, y) = f_1(x) \oplus f_2(y)$ is a $(k_1 + k_2 + 1)$th-order correlation immune function on $V_{n_1+n_2}$ whose nonlinearity satisfies*

$$N_g \geqq d_1 2^{n_2} + d_2 2^{n_1} - 2d_1 d_2,$$

where $x = (x_1, x_2, \ldots, x_{n_1})$ and $y = (y_1, y_2, \ldots, y_{n_2})$. In particular g is balanced if either f_1 or f_2 is balanced.

6.3 Combination of Four Correlation Immune Functions

This section show that from four correlation immune functions, we can obtain a new functions that achieves a higher order of correlation immunity.

Theorem 8 *Let f_1 and f_2 be pth-order correlation immune functions on V_m, and let h_1 and h_2 be qth-order correlation immune functions on V_n. Let ξ_1, ξ_2, η_1 and η_2 be the sequences of f_1, f_2, h_1 and h_2 respectively. Let ζ be a $(1, -1)$-sequence obtained from ξ_1, ξ_2, η_1 and η_2 in the following way:*

$$\zeta = \frac{1}{2}(\xi_1 + \xi_2) \otimes \eta_1 + \frac{1}{2}(\xi_1 - \xi_2) \otimes \eta_2 \tag{16}$$

where $+$ denotes the component-wise integer addition and $\otimes$ denotes the Kronecker product. Then the function corresponding to ζ is a $(p + q + 1)$th-order correlation immune function on V_{m+n}.

Proof. Similarly to the proof of Lemma 7, we let φ be a linear function on V_{m+n} and L be the sequence of φ. By Lemma 2, L is a row of H_{m+n}. In addition, φ can be written as $\varphi = \langle \gamma, z \rangle = \langle \alpha, x \rangle \oplus \langle \beta, y \rangle$, where $\gamma = (\alpha, \beta) \in V_{m+n}$, $\alpha \in V_m$, $\beta \in V_n$, $z = (x_1, \ldots, x_m, y_1 \ldots, y_n)$, $x = (x_1, x_2, \ldots, x_m)$ and $y = (y_1, y_2, \ldots, y_n)$. Since $H_{m+n} = H_m \otimes H_n$ L can be expressed as $L = \ell_1 \otimes \ell_2$, where ℓ_1 is a row of H_m, and ℓ_2 is a row of H_n. By the same reasoning as in the proof of Lemma 7, it can be shown that ℓ_1 is the sequence of $\langle \alpha, x \rangle$, and ℓ_2 is the sequence of $\langle \beta, y \rangle$. Thus we have

$$\begin{aligned} \langle \zeta, L \rangle &= \frac{1}{2}\langle (\xi_1 + \xi_2) \otimes \eta_1, \ell_1 \otimes \ell_2 \rangle + \frac{1}{2}\langle (\xi_1 - \xi_2) \otimes \eta_2, \ell_1 \otimes \ell_2 \rangle \\ &= \frac{1}{2}\langle (\xi_1 + \xi_2), \ell_1 \rangle \langle \eta_1, \ell_2 \rangle + \frac{1}{2}\langle (\xi_1 - \xi_2), \ell_1 \rangle \langle \eta_2, \ell_2 \rangle. \end{aligned} \tag{17}$$

For $\gamma \in V_{m+n}$ with $W(\gamma) \leqq p + q + 1$, we have either $W(\alpha) \leqq p$ or $W(\beta) \leqq q$. This implies that either of the following two situations occurs: (1) $\langle \xi_1, \ell_1 \rangle = 0$ and $\langle \xi_2, \ell_1 \rangle = 0$, and (2) $\langle \eta_1, \ell_2 \rangle = 0$ and $\langle \eta_2, \ell_2 \rangle = 0$. As a consequence, we have $\langle \zeta, L \rangle = 0$. □

Note that a similar technique to the construction (16) has been used in obtaining higher order Hadamard matrices from lower order Hadamard matrices [6].

7 Conclusion

We have studied correlation immune functions using the theory of Hadamard matrices. In particular, we have presented a new method for directly constructing correlation immune functions. It is shown that the method generates the same set of functions as that by a method of Camion et al. The new method is more convenient for use in practice since it allows one to calculate the nonlinearity of functions obtained and to discuss the algebraic degrees and propagation characteristics of the functions. Three methods for obtaining correlation immune functions on a higher dimensional space from known correlation immune functions on a lower dimensional space are also presented. We believe that these various methods of generating correlation immune functions, by direct construction or by combining known correlation immune functions, will find a wide range of applications in computer security.

References

[1] W. Blaser and P. Heinzmann. New cryptographic device with high security using public key distribution. In *Proceedings of IEEE Student Paper Contest 1979-1980*, pages 145–153, 1982.

[2] P. Camion, C. Carlet, P. Charpin, and N. Sendrier. On correlation-immune functions. In *Advances in Cryptology - CRYPTO'91*, volume 576, Lecture Notes in Computer Science, pages 87–100. Springer-Verlag, Berlin, Heidelberg, New York, 1991.

[3] Xiao Guo-zhen and J. L. Massey. A spectral characterization of correlation-immune combining functions. *IEEE Transactions on Information Theory*, 34 No. 3:569–571, 1988.

[4] R. A. Rueppel. *Analysis and Design of Stream Ciphers*. Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Berlin, Heidelberg, New York, London, Paris, Tokyo, 1986. In Communications and Control Engineering Series, Editors: A. Fettweis, J. L. Massey and M. Thoma.

[5] R. A. Rueppel. Stream ciphers. In G. J. Simmons, editor, *Contemporary Cryptography: the Science of Information Integrity*, chapter 2, pages 65–134. IEEE Press, New York, 1992.

[6] J. Seberry and M. Yamada. Hadamard matrices, sequences, and block designs. In J. H. Dinitz and D. R. Stinson, editors, *Contemporary Design Theory: A Collection of Surveys*, chapter 11, pages 431–559. John Wiley & Sons, Inc, 1992.

[7] J. Seberry and X. M. Zhang. Highly nonlinear 0-1 balanced functions satisfying strict avalanche criterion. In *Advances in Cryptology - AUSCRYPT'92*. Springer-Verlag, Berlin, Heidelberg, New York, 1993. to appear.

[8] T. Siegenthaler. Correlation-immunity of nonlinear combining functions for cryptographic applications. *IEEE Transactions on Information Theory*, IT-30 No. 5:776–779, 1984.

[9] W. D. Wallis, A. Penfold Street, and J. Seberry Wallis. *Combinatorics: Room Squares, sum-free sets, Hadamard Matrices*, volume 292 of Lecture Notes in Mathematics. Springer-Verlag, Berlin, Heidelberg, New York, 1972.