Three Characterizations of Non-binary Correlation-Immune and Resilient Functions

K. Gopalakrishnan
Department of Computer Science and Engineering
University of Nebraska - Lincoln
Lincoln, NE 68588

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

January 16, 1997

Abstract

A function $f(X_1, X_2, \ldots, X_n)$ is said to be t *th-order correlation-immune* if the random variable $Z = f(X_1, X_2, \ldots, X_n)$ is independent of every set of t random variables chosen from the independent equiprobable random variables $X_1, X_2, \ldots, X_n$. Additionally, if all possible outputs are equally likely, then f is called a $t-resilient$ function. In this paper, we provide three different characterizations of t th-order correlation immune functions and resilient functions where the random variable is over $GF(q)$. The first is in terms of the structure of a certain associated matrix. The second characterization involves Fourier transforms. The third characterization establishes the equivalence of resilient functions and large sets of orthogonal arrays.

keywords: Correlation-immune functions, resilient functions, stream ciphers, Fourier transforms, orthogonal arrays.

1 Definitions

Let $GF(q)$ denote the Galois Field with q elements, where $q = p^a$ is a prime power. Let $f : [GF(q)]^n \longrightarrow [GF(q)]^m$ be a function and let $\{X_1, X_2, \ldots, X_n\}$ be the set of random input variables assuming values from $GF(q)$ with independent equiprobable distributions (that is, every possible input vector occurs with equal probability $1/q^n$). Such a function is said to have *balanced input*. Let $\{Z_1, Z_2, \ldots, Z_m\}$ denote the set of output variables of the function f. The function f is said to be *correlation-immune* with respect to $T \subseteq \{1, 2, \ldots, n\}$ (or equivalently *independent* of T) if the probability distribution of the random variable $f(X_1, X_2, \ldots, X_n)$ is unaltered when $\{X_i : i \notin T\}$ is a set of independent equiprobable random variables and $\{X_i : i \in T\}$ is a set of constants. f is said to be t *th-order correlation-immune* if for every $T \subseteq \{1, 2, \ldots, n\}$ of cardinality at most t, the function f is correlation-immune with respect to T. More formally, the property can be stated as follows: For every t-subset $\{i_1, \ldots, i_t\} \subseteq \{1, 2, \ldots, n\}$, for every choice of $z_j \in GF(q)(1 \le j \le t)$, and for every $(y_1, \ldots, y_m) \in [GF(q)]^m$, we have

$$Pr(f(x_1, \ldots, x_n) = (y_1, \ldots, y_m) \mid x_{i_j} = z_j, 1 \le j \le t)$$
$$= Pr(f(x_1, \ldots, x_n) = (y_1, \ldots, y_m)).$$

The function f is said to have *balanced output* if every possible output m-tuple occurs with equal probability $1/q^m$. f is said to be t*-resilient* if it is t th-order correlation-immune and has balanced output. More formally, the function f is said to be $t - resilient$, if for every subset $\{i_1, \ldots, i_t\} \subseteq \{1, \ldots, n\}$, for every choice of $z_j \in GF(q)(1 \le j \le t)$, and for every $(y_1, \ldots, y_m) \in [GF(q)]^m$, we have

$$Pr(f(x_1, \ldots, x_n) = (y_1, \ldots, y_m) \mid x_{i_j} = z_j, 1 \le j \le t)$$
$$= Pr(f(x_1, \ldots, x_n) = (y_1, \ldots, y_m)) = \frac{1}{q^m}.$$

Correlation-immune functions were defined by Siegenthaler in [8] and further studied in [6], [7] and [3]. In stream ciphers the running key generator used typically involves a memoryless combining function. In order to increase the security of the cipher system, the combining function must be selected with care. The general approach is to select the function so as to increase the linear complexity of the resulting key stream. However, even if the linear complexity of the key stream is high, sometimes the system can be broken by a correlation attack [7]. In order to prevent this, the function selected should be correlation-immune to a desired degree.

The concept of binary resilient functions was introduced by Chor *et al* [4] and Bennett *et al* [2]. Two possible applications of resilient functions were mentioned in [2] and [4]. The first application concerns the generation of shared random strings in the presence of faulty processors. The second involves renewing a partially leaked cryptographic key.

In the binary case (i.e., when $q = 2$), a characterization of correlation-immune functions was obtained by Zhen and Massey [6] in terms of the Walsh transform. It was shown [6] that a Boolean combining function $f(x)$ of n variables is t-th order correlation-immune if and only if its Walsh transform $F(u)$ vanishes for all u with Hamming weights between 1 and t, inclusive. Essentially the same characterization (for resilient functions) is obtained by Chor et al [4].

In [3], binary correlation-immune functions with $m = 1$ were characterized in terms of orthogonal arrays. A similar characterization for binary resilient functions (for arbitrary m) was proved by Stinson in [9].

In this paper, we provide three different characterizations of correlation-immune and resilient functions where the random variable is over $GF(q)$. The rest of the paper is organized as follows. Section 2 states and proves a key lemma viz., *linear combination lemma.* This lemma is a generalization of the so-called XOR-lemma that applies in the binary case. We use it in the proof of the matrix characterization given later, but it also can be used to extend the matrix and Fourier transform characterizations from $m = 1$ to $m > 1$. In Section 3, we give a characterization of correlation-immunity in terms of the structure of certain associated matrices. In Section 4, an even more succinct characterization of correlation-immunity is obtained using Fourier transforms. This generalizes the Walsh transform characterization mentioned above. Finally in Section 5, we present a characterization of correlation-immune and resilient functions in terms of orthogonal arrays. This generalizes the characterizations in [3] and [9].

2 Linear Combination Lemma

In this section, we prove an important lemma, which we call the "linear combination lemma." The linear combination lemma is a generalization of the *XOR-lemma* proved independently in [6] and [4] for the binary case. As we shall see, our generalization depends on certain properties of finite affine planes.

Lemma 2.1 (Linear Combination Lemma) *The discrete random variable Z is independent of the k independent random variables $Y_1, Y_2, \ldots Y_k$ if and only if Z is independent of the sum $\lambda_1 Y_1 + \lambda_2 Y_2 + \ldots + \lambda_k Y_k$ for every choice of $\lambda_1, \lambda_2 \ldots \lambda_k$, not all zeroes, in $GF(q)$ (i.e., iff Z is independent of any non-trivial linear combination of the k variables).*

Proof: The necessity of the abovementioned condition is obvious and it remains only to show the sufficiency. Also sufficiency is trivial for $k = 1$. Let us consider the case $k = 2$. Suppose that Z is independent of any linear combination of Y_1 and Y_2. The following proof applies to any possible value z of Z. We need the following notations. Let $\alpha_i = Pr(Y_1 = i)$ and $\beta_i = Pr(Y_2 = i)$ where $i \in GF(q)$. Let $x_{i,j} = Pr(Y_1 = i, Y_2 = j \mid Z = z)$ where $i, j \in GF(q)$. Thus there are q^2 different variables $x_{i,j}$. However, the following relations hold among these q^2 different variables. The most obvious one is that

$$\sum_{i \in GF(q)} \sum_{j \in GF(q)} x_{i,j} = 1.$$

Consider a particular linear combination of Y_1 and Y_2, say $\lambda_1 Y_1 + \lambda_2 Y_2$. Let $\mu_j(i)$ denote the unique value of Y_2 such that $\lambda_1 i + \lambda_2 Y_2 = j$. Since Z is independent of any linear combination of Y_1 and Y_2, it follows that the following relation holds among the q^2 variables for each value of $j \in GF(q)$.

$$\begin{aligned}
\sum_{i \in GF(q)} x_{i,\mu_j(i)} &= Pr(\lambda_1 Y_1 + \lambda_2 Y_2 = j \mid Z = z) \\
&= Pr(\lambda_1 Y_1 + \lambda_2 Y_2 = j) \\
&= \sum_{i \in GF(q)} Pr(Y_1 = i) Pr(Y_2 = \mu_j(i)) \\
&= \sum_{i \in GF(q)} \alpha_i \beta_{\mu_j(i)}.
\end{aligned}$$

Thus q relations that hold among the q^2 variables are obtained from one particular linear combination. There are $q^2 - 1$ different non-trivial linear combinations as λ_1 and λ_2 can take any value from $GF(q)$ except that they cannot simultanously be zero. Call a linear combination l_1 equivalent to another linear combination l_2 iff l_2 is a non-zero scalar multiple of l_1. It is easy to check that this is indeed an equivalence relation on the set of non-trivial linear combinations and that the number of different equivalence classes is

$$\frac{q^2-1}{q-1} = q+1.$$

From each one of these equivalence classes, we get q relations as explained above, yielding a total of $(q+1)(q) = q^2 + q$ relations.

Let $Ax = b$ be the matrix form of the above set of $q^2 + q$ equations, where x is a column vector of size q^2, b is a column vector of size $q^2 + q$ and A is a $(q^2+q) \times q^2$ matrix. It is clear that A is a 0-1 matrix. Infact, A is simply the incidence matrix of the finite affine plane of order q, or equivalently, that of a BIBD (balanced incomplete block design) with parameters $(q^2, q^2+q, q+1, q, 1)$. It is well known that the blocks of a BIBD with parameters (v, b, r, k, λ) span the real vector space of dimension v. Hence the row rank of the A matrix is q^2. Thus there are q^2 linearly independent equations that hold among the q^2 variables $x_{i,j}$ and hence the above system of equations have at most one solution. It is a routine matter to check that these equations indeed have the unique solution

$$Pr(Y_1 = i, Y_2 = j \mid Z = z) = x_{i,j} = \alpha_i \times \beta_j = Pr(Y_1 = i, Y_2 = j)$$

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. Thus the lemma is established for $k = 2$.

Now suppose that the sufficiency condition has been established for $k < K$. We will show that the sufficiency condition also holds for $k = K$ and hence by the principle of mathematical induction, the lemma is true. Suppose that Z is independent of the sum $\lambda_1 Y_1 + \lambda_2 Y_2 + \ldots + \lambda_K Y_K$ for all $\lambda_1, \lambda_2, \ldots \lambda_K$, not all zero in GF(q). Choosing $\lambda_1 = \lambda_2 = 0$ implies, by the induction hypothesis, that $Y_3, \ldots Y_K, Z$ are independent. Let $\mu_1 Y_1 + \mu_2 Y_2$ be an arbitrary nontrivial linear combination of Y_1 and Y_2. Treat this linear combination itself as a single random variable. It is clear that Z is independent of the sum

$$\lambda(\mu_1 Y_1 + \mu_2 Y_2) + \lambda_3 Y_3 + \ldots + \lambda_K Y_K.$$

By the induction hypothesis, it follows that $\mu_1 Y_1 + \mu_2 Y_2$ is independent of $Y_3, Y_4, \ldots, Y_K, Z$. Thus for every choice of $y_3, y_4, \ldots, y_K, z$, the previous equations still apply after replacing $Pr(Y_1 = i, Y_2 = j \mid Z = z)$ by

$$Pr(Y_1 = i, Y_2 = j \mid Y_3 = y_3, Y_4 = y_4, \ldots, Y_K = y_k, Z = z)$$

everywhere. Thus it follows that the pair Y_1, Y_2 is independent of $Y_3, Y_4, \ldots Y_K, Z$ and hence the random variables $Y_1, Y_2, \ldots Y_K, Z$ are all independent. $\square$

3 Matrix Characterization

In this section, a characterization of t th order correlation-immune functions is developed in terms of certain matrices. Let $X = (X_1, X_2, \ldots X_n)$ and $u = (u_1, u_2, \ldots u_n)$ be n-tuples over $GF(q)$, and define their *dot product* as

$$X \circ u = X_1 u_1 + X_2 u_2 + \ldots + X_n u_n \in GF(q).$$

Let $f : [GF(q)]^n \longrightarrow GF(q)$ be a t th order correlation-immune function. For a fixed n-tuple u, construct a $q \times q$ real matrix $N = N(u) = (\eta_{i,j})$, where

$$\eta_{i,j} = q^n \cdot Pr(X \circ u = i \text{ and } f(X) = j),$$

$i, j \in GF(q)$. Further, define

$$\beta_i = \sum_{j \in GF(q)} \eta_{i,j} = q^n \cdot Pr(X \circ u = i)$$

and

$$\alpha_j = \sum_{i \in GF(q)} \eta_{i,j} = q^n \cdot Pr(f(X) = j).$$

Let $W(u)$ denote the Hamming weight of u, i.e. the number of non-zero components in u. Then for any $u \neq 0$, $X \circ u$ is a linear combination of $W(u)$ of the random variables $X_1, X_2, \ldots, X_n$. It follows from the linear combination lemma that if f is tth order correlation-immune, then $Z = f(X)$ is independent of every random variable $u \circ X$ for which $1 \leq W(u) \leq t$. Hence, we have the following:

$$\begin{aligned} \alpha_j &= q^n \cdot Pr(f(X) = j) \\ &= q^n \cdot Pr(f(X) = j \mid X \circ u = i) \\ &= q^n \frac{Pr(f(X) = j \text{ and } X \circ u = i)}{Pr(X \circ u = i)} \\ &= \frac{q^n \eta_{i,j}}{\beta_i}. \end{aligned}$$

Hence, for all i, j, we have

$$\alpha_j \beta_i = q^n \eta_{i,j}. \tag{1}$$

So far, we have not used the fact that f has balanced input. Since the input is balanced, β_i is independent of i. Since $\sum \beta_i = q^n$, it follows that $\beta_i = q^{n-1}$ for all i. Thus, from Equation (1), we have that $\alpha_j = q\eta_{i,j}$ for all i, j. This says that all the rows of the N matrix are identical.

All of these steps can be reversed, so we obtain the following theorem.

Theorem 3.1 *The function $f : [GF(q)]^n \longrightarrow GF(q)$ is correlation-immune of order t if and only if the rows of the matrix $N(u)$ are all identical for any u such that $1 \leq W(u) \leq t$.*

If we further assume that the output is balanced, then α_j is independent of j. Since $\sum \alpha_j = q^n$, it follows that $\alpha_j = q^{n-1}$ for all j. Then, by Theorem 3.1, it must be the case that every entry of $N(u)$ is q^{n-2}. We have proved the following.

Theorem 3.2 *Let J denote the $q \times q$ matrix in which all the entries are equal to 1. The function $f : [GF(q)]^n \longrightarrow GF(q)$ is t-resilient if and only if*

$$N(u) = q^{n-2} J$$

for any u such that $1 \leq W(u) \leq t$.

4 Fourier Transform Characterization

First, a brief introduction to the Fourier transform is presented, which will be used later for characterizing t-resilient functions. More details on Fourier transform can be found in standard texts [1].

Let ω be a principal pth root of unity in the field of complex numbers. Let $M = (m_{i,j})$ be the $p \times p$ matrix such that $m_{i,j} = \omega^{ij}$, for $0 \leq i, j \leq p-1$. The matrix M is non-singular and thus M^{-1} exists. Moreover, it has the simple form given by $m_{i,j}^{-1} = (\frac{1}{p})\omega^{-ij}$, for $0 \leq i, j \leq p-1$. Let $a = (a_0, a_1, \ldots, a_{p-1})$ be a row vector with p elements. The vector $b = F(a) = aM$ whose ith component b_i, $0 \leq i < p$, is

$$\sum_{k=0}^{p-1} a_k \omega^{ik},$$

is called the *Fourier transform* of a. The vector $F^{-1}(a) = aM^{-1}$ whose ith component, $0 \leq i \leq p-1$, is

$$\frac{1}{p} \sum_{k=0}^{p-1} a_k \omega^{-ik},$$

is the *inverse Fourier transform* of a. Clearly, the inverse transform of the transform of a is a itself.

Computing the Fourier transform of a vector $(a_0, a_1, \ldots, a_{p-1})$ is equivalent to converting the coefficient representation of the polynomial

$$\sum_{i=0}^{p-1} a_i x^i$$

to its value representation at $\omega^0, \omega^1, \ldots, \omega^{p-1}$. Similarly, the inverse Fourier transform is equivalent to interpolating a polynomial given its values at the pth roots of unity.

Let $f : [GF(p)]^n \longrightarrow GF(p)$ be a tth order correlation-immune function. We will obtain a succinct characterization of such functions using Fourier transforms and the matrix characterization described above. Note that we are restricting our attention here to functions over $GF(p)$, where p is a prime number, for the sake of simplicity. The extension of this characterization to functions over $GF(q)$, where q is a prime power, is outlined at the end of the section.

Consider a particular n-tuple $u = (u_1, u_2, \ldots, u_n)$ over $GF(p)$ with $1 \leq W(u) \leq t$. Let N be the matrix as defined in Section 3 and let M be the matrix as defined above. In Section 3, it was observed that, the function f is tth order correlation-immune if and only if the rows of the N matrix are identical to each other for $1 \leq W(u) \leq t$.

Let $S = N \times M$ be a $p \times p$ matrix. It is clear that

$$s_{i,j} = \sum_{k=0}^{p-1} \eta_{i,k} m_{k,j} = \sum_{k=0}^{p-1} \eta_{i,k} \omega^{kj}.$$

If one thinks of a row of N as the representation of a $(p-1)$st degree polynomial by coefficients, then the corresponding row of the S matrix is a representation of the same polynomial by its values at $\omega^0, \omega^1, \ldots, \omega^{p-1}$. Thus the rows of N are identical to each other if and only if the rows of S are identical to each other.

Theorem 4.1 *A function $f : [GF(p)]^n \longrightarrow GF(p)$ is t th order correlation-immune if and only if*

$$\sum_x \alpha^{x \circ u} \beta^{f(x)} = 0 \tag{2}$$

for all α and β which are pth roots of unity, $\alpha \neq 1$, and for all $u \in [GF(p)]^n$ such that $1 \leq W(u) \leq t$. [Note that the sum is over all x in $[GF(p)]^n$ and that both $f(x)$ and $x \circ u$ in the exponent must be treated as an integer between 0 and $p-1$ rather than as an element of $GF(p)$ in the above expression.]

Proof: First we shall prove that if Equation (2) is satisfied then f is tth order correlation-immune. Since the pth roots of unity form a cyclic group and ω is a principal pth root of unity, we have $\alpha = \omega^a$ and $\beta = \omega^b$ where $1 \leq a \leq p-1$ and $0 \leq b \leq p-1$.

Then

$$\begin{aligned}
0 &= \sum_x \alpha^{x \circ u} \beta^{f(x)} \\
&= \sum_x (\omega^a)^{x \circ u} (\omega^b)^{f(x)} \\
&= \sum_{i=0}^{p-1} \sum_{j=0}^{p-1} \eta_{i,j} \omega^{ai} \omega^{bj} \\
&= \sum_{i=0}^{p-1} \omega^{ai} \sum_{j=0}^{p-1} \eta_{i,j} \omega^{bj} \\
&= \sum_{i=0}^{p-1} \omega^{ai} s_{i,b}.
\end{aligned}$$

Let $T = M \times S = (t_{a,b})$. ¿From the above computations, we see that $t_{a,b} = 0$ if $a \neq 0$. Further, $t_{0,b} = c_b$, where c_b is the sum of the entries in column b of S.

Hence, the structure of the T matrix is completely known. Since we have $T = M \times S$, we can solve for S by premultiplying by M^{-1}. Thus $S = M^{-1} \times T$. Since the first column of M^{-1} has all entries $1/p$ and the only non-zero entries of the T matrix lie in the first row, it is clear that

$$s_{i,j} = \frac{c_j}{p}$$

for all $0 \leq j \leq p-1$. Thus the entries in each column of the S matrix are the same. So the rows of the S matrix are identical to each other. Hence it follows that f is tth order correlation-immune.

Conversely, if f is tth order correlation-immune, then the rows the S matrix are identical to each other by virtue of the matrix characterization. In other words the entries in each column of the S matrix are the same. Since ω is a principal pth root of unity it follows that

$$\sum_{j=0}^{p-1} \omega^{ij} = 0$$

for $1 \leq i \leq p-1$. i.e., the sum of every row of the M matrix is zero, except for the first row. Define $T = M \times S$. It follows that $t_{a,b} = 0$ for $1 \leq a \leq p-1$ and $0 \leq b \leq p-1$. It is easy to check that such a condition on the T matrix implies satisfaction of Equation (2). $\Box$

Although the above theorem concerns functions over $GF(p)$, where p is a prime, it can be easily extended to functions over $GF(q)$, q a prime power. This extension is due to Bierbrauer (private communication).

Let $q = p^a$. We define a mapping $tr : GF(q) \rightarrow GF(p)$, which is called the *trace function* as follows. If $\mu \in GF(q)$, then

$$Tr(\mu) = \mu + \mu^p + \mu^{p^2} + \ldots + \mu^{p^{a-1}}.$$

As before, let ω be a principal pth root of unity in the field of complex numbers. Let M be the $q \times q$ matrix, indexed by pairs (x, y) of elements of $GF(q)$, defined by $m_{x,y} = \omega^{tr(xy)}$. Using standard properties of the trace function, the following Theorem on the characterization of correlation-immune functions over $GF(q)$ can be proved.

Theorem 4.2 *A function $f : [GF(q)]^n \longrightarrow GF(q)$ (where $q = p^a$ is a prime power) is t th order correlation-immune if and only if*

$$\sum_{x \in [GF(q)]^n} \omega^{tr(v \cdot (x \circ u) + w \cdot f(x))} = 0.$$

for all $(v, w) \in [GF(q)]^2, v \neq 0$ and for all $u \in [GF(q)]^n$ such that $1 \leq W(u) \leq t$.

The proof is analogous to that of Theorem 4.1 and hence is omitted.

The matrix and Fourier transform characterizations were proved for the case $m = 1$. To extend these characterizations to $m > 1$, it suffices to use the linear combination lemma. That is, every linear combination of the m output varibles should satisfy the relevant characterization with $m = 1$.

5 Orthogonal Array Characterization

Correlation-immune functions and resilient functions are closely related to orthogonal arrays. These connections were explored and several results were proved for the binary case in [9], [4] and [3]. This section generalizes those results to the non-binary case.

An *orthogonal array* $OA_\lambda(t,k,v)$ is a $\lambda v^t \times k$ array of v symbols, such that in any t columns of the array every one of the possible v^t ordered t-tuples of symbols occurs in exactly λ rows. An orthogonal array is said to be *simple* if no two rows are identical.

A *generalized large set of orthogonal arrays* $GLOA[\mu]_{\lambda_1,\ldots,\lambda_\mu}(t,k,v)$ is defined to be a collection of μ simple arrays $OA_{\lambda_i}(t,k,v)$, $1 \le i \le \mu$ such that every possible k-tuple of symbols occurs in exactly one of the orthogonal arrays in the collection. Equivalently, the union of the orthogonal arrays forms an $OA_1(k,k,v)$. Clearly, it must be the case that

$$\sum_{i=1}^{\mu} \lambda_i = v^{k-t}.$$

A *large set of orthogonal arrays* $LOA_\lambda(t,k,v)$ is simply a special case of $GLOA[\mu]_{\lambda_1,\ldots,\lambda_\mu}(t,k,v)$ in which

$$\lambda_1 = \lambda_2 = \ldots = \lambda_\mu = \lambda.$$

In other words, a large set of orthogonal arrays $LOA_\lambda(t,k,v)$ is a set of v^{k-t}/λ simple arrays $OA_\lambda(t,k,v)$ such that every possible k-tuple of symbols occurs in exactly one of the orthogonal arrays in the set.

We will refer to a t th order correlation-immune function with balanced input $f : [GF(q)]^n \longrightarrow [GF(q)]^m$ (where q is the power of a prime) as an (n,m,t,q) - CI function. Similarly we will refer to a t-resilient function $f : [GF(q)]^n \longrightarrow [GF(q)]^m$ as an (n,m,t,q) - resilient function.

The following theorem elucidates the connection between correlation-immune functions and orthogonal arrays.

Theorem 5.1 *An (n,m,t,q) - CI function is equivalent to a generalized large set of orthogonal arrays $GLOA[q^m]_{\lambda_1,\ldots,\lambda_{q^m}}(t,n,q)$.*

Proof: First, suppose $f : [GF(q)]^n \longrightarrow [GF(q)]^m$ is an (n,m,t,q) - resilient function. For any $y \in [GF(q)]^m$, form an array A_y whose rows are the vectors in the inverse image of $f^{-1}(y)$. A_y is a $|f^{-1}(y)| \times n$ array with symbols from $GF(q)$. It is clear that the q^m arrays A_y together contain every possible n-tuple as a row, so if each A_y is an orthogonal array of type $OA(t,n,q)$, then we automatically get a generalized large set of orthogonal arrays $GLOA[q^m]_{\lambda_1,\ldots,\lambda_{q^m}}(t,n,q)$.

Let $\{i_1, i_2, \ldots, i_t\} \subseteq \{1,2,\ldots,n\}$ be a t-subset, and let $z_j \in GF(q)(1 \le j \le t)$. For every $y \in [GF(q)]^m$, let λ_y denote the number of rows in A_y in which z_j occurs in column i_j for $1 \le j \le t$. It is easy to see that

$$\sum_{y \in [GF(q)]^m} \lambda_y = q^{n-t}.$$

Since the input is balanced,

$$Pr(f(x_1,\ldots,x_n) = (y_1,\ldots,y_m) \mid x_{i_j} = z_j, 1 \le j \le t) = \frac{\lambda_y}{q^{n-t}}.$$

Since f is t-th order correlation-immune, we get

$$Pr(f(x_1, \ldots, x_n) = (y_1, \ldots, y_m) \mid x_{i_j} = z_j, 1 \leq j \leq t)$$
$$= Pr(f(x_1, \ldots, x_n) = (y_1, \ldots, y_m)).$$

These two equations together imply that

$$\lambda_y = q^{n-t} \times Pr(f(x_1, \ldots, x_n) = (y_1, \ldots, y_m)).$$

This shows that λ_y is independent of the particular choices made for $i_1, i_2, \ldots i_t$ and $z_j (1 \leq j \leq t)$ and thus it follows that each A_y is an orthogonal array $OA_{\lambda_y}(t, n, q)$, as desired.

Conversely, suppose we start with a $GLOA[q^m]_{\lambda_1, \ldots, \lambda_{q^m}}(t, n, q)$. There are q^m arrays in the large set; name them $A_y, y \in [GF(q)]^m$. Then define a function $f : [GF(q)]^n \longrightarrow [GF(q)]^m$ by the rule

$$f(x_1, \ldots, x_n) = (y_1, \ldots, y_m) \Longleftrightarrow (x_1, \ldots, x_n) \in A_{(y_1, \ldots, y_m)}.$$

Now

$$Pr(f(x_1, \ldots, x_n) = (y_1, \ldots, y_m) \mid x_{i_j} = z_j, 1 \leq j \leq t)$$

$$\begin{aligned} &= \frac{\lambda_y}{q^{n-t}} \\ &= \frac{|A_y|}{q^n} \\ &= Pr(f(x_1, \ldots, x_n) = (y_1, \ldots, y_m)). \end{aligned}$$

Hence f is an (n, m, t, q) - CI function. ▯

Note: Theorem 5.1 was proved in the binary case for $m = 1$ by Camion *et al* in [3]. Their proof uses sophisticated techniques developed by Delsarte in [5], as well as the Walsh transform characterization. Our proof is a direct proof, and is more elementary and combinatorial in nature.

Theorem 5.2 *An (n, m, t, q) - resilient function is equivalent to a large set of orthogonal arrays $LOA_{q^{n-m-t}}(t, n, q)$.*

Proof: The proof is analogous to that of Theorem 5.1 and hence some details are omitted. If we start with (n, m, t, q) - resilient function, then, as before

$$\lambda_y = q^{n-t} \times Pr(f(x_1, \ldots, x_n) = (y_1, \ldots, y_m)).$$

Since the output is balanced,

$$Pr(f(x_1, \ldots, x_n) = (y_1, \ldots, y_m)) = q^{-m}.$$

Hence, $\lambda_y = q^{n-m-t}$ for all y, and thus it follows that the A_y's form an $LOA_{q^{n-m-t}}(t, n, q)$, as desired.

The converse is also straightforward. ▯

Remark: It was shown in [2] and [4] that resilient functions can be constructed from linear codes. Connections between resilient functions, orthogonal arrays and error-correcting codes were studied in the binary case in [9]. Almost all the results found in [9] immediately generalize to functions over $GF(q)$ and are omitted from the discussion here for the sake of brevity.

Acknowledgements

This research was supported by NSF grant CCR-9121051. We would like to thank Juergen Bierbrauer and Wen-ai Jackson for several helpful discussions.

References

[1] Alfred V. Aho, John E. Hopcroft, and Jeffrey D. Ullman. *The Design and Analysis of Computer Algorithms.* Computer Science and Information Processing. Addison-Wesley Publishing Company, 1975.

[2] C. H. Bennett, G. Brassard, and J. M. Robert. Privacy amplification by public discussion. *SIAM J. Computing*, **17** (1988), 210–229.

[3] P. Camion, C. Carlet, P. Charpin, and N. Sendrier. On correlation-immune functions. *Lecture Notes in Computer Science*, **576** (1992), 86–100 (Advances in Cryptology - CRYPTO '91 Proceedings).

[4] Benny Chor, Oded Goldreich, Johan Håstad, Joel Friedman, Steven Rudich, and Roman Smolensky. The bit extraction problem or t-resilient functions. *26th IEEE Symposium on Foundations of Computer Science*, 1985, pages 396–407.

[5] P. Delsarte. An algebraic approach to the association schemes of coding theory. *Philips Research Reports Supplements*, **10**, 1973. Thesis, Universite Catholique de Louvain.

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

[7] Rainer A. Rueppel. *Analysis and Design of Stream Ciphers.* Springer-Verlag, 1986.

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

[9] D. R. Stinson. Resilient functions and large sets of orthogonal arrays. *Congressus Numer.*, **92** (1993), 105–110.