Resilient functions and large sets of orthogonal arrays

D. R. Stinson
Computer Science and Engineering Department
and Center for Communication and Information Science
University of Nebraska
Lincoln, NE 68588-0115, U.S.A.
stinson@bibd.unl.edu

Abstract

In this paper we discuss the connections between resilient functions, large sets of orthogonal arrays and error-correcting codes. Some recent results on resilient functions are then derived as consequences of known results on orthogonal arrays from design theory.

1 Introduction

The concept of resilient functions was introduced independently in the two papers Chor *et al* [4] and Bennett, Brassard and Robert [1]. Here is the definition. Let $n \geq m \geq 1$ be integers and suppose

$$f : \{0,1\}^n \rightarrow \{0,1\}^m.$$

We will think of f as being a function that accepts n input bits and produces m output bits. Let $t \leq n$ be an integer. Suppose $(x_1, \ldots, x_n) \in \{0,1\}^n$, where the values of t arbitrary input bits are fixed by an opponent, and the remaining $n-t$ input bits are chosen independently at random. Then f is said to be $t-$*resilient* provided that every possible output $m-$tuple is equally likely to occur. More formally, the property can be stated as follows: For every $t-$subset $\{i_1, \ldots, i_t\} \subseteq \{1, \ldots, n\}$, for every choice of $z_j \in \{0,1\}$ $(1 \leq j \leq t)$, and for every $(y_1, \ldots, y_m) \in \{0,1\}^m$, we have

$$p(f(x_1, \ldots, x_n) = (y_1, \ldots, y_m) | x_{i_j} = z_j, 1 \leq j \leq t) = \frac{1}{2^m}.$$

We will refer to such a function f as an $(n, m, t)-$resilient function.

A closely related concept is that of a correlation-immune function, which is defined by Siegenthaler in [11] and further studied in [10], [6] and [3]. Let $n \geq 1$ be an integer and suppose $f : \{0,1\}^n \rightarrow \{0,1\}$. As before, suppose $(x_1, \ldots, x_n) \in \{0,1\}^n$, where the values of t arbitrary input bits are fixed by an opponent, and the remaining $n-t$ input bits are chosen independently at random. Then f is said to be *correlation-immune of order* t provided that for every $t-$subset $\{i_1, \ldots, i_t\} \subseteq \{1, \ldots, n\}$, for every choice of $z_j \in \{0,1\}$ $(1 \leq j \leq t)$, and for $y = 0, 1$, we have

CONGRESSUS NUMERANTIUM 92(1993), pp.105–110

$$p(f(x_1,\ldots,x_n) = y | x_{i_j} = z_j, 1 \le j \le t) = p(f(x_1,\ldots,x_n) = y).$$

A correlation-immune function is *balanced* if

$$p(f(x_1,\ldots,x_n) = y | x_{i_j} = z_j, 1 \le j \le t) = 1/2).$$

In other words, a balanced correlation-immune function is the same thing as an $(n, 1, t)$−resilient function.

Two possible applications of resilient functions are mentioned in [1] 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. Correlation-immune functions are used in stream ciphers as combining functions for running-key generators that are resistant to a correlation attack (see, for example, Rueppel [10]).

Many interesting results on resilient functions can be found in [1] and [4]. The basic problem is to maximize t given m and n; or equivalently, to maximize m given n and t. Here are some examples from [4] (all addition is modulo 2):

(1) $m = 1$, $t = n - 1$. Define $f(x_1,\ldots,x_n) = x_1 + \ldots + x_n$.

(2) $m = n - 1$, $t = 1$. Define $f(x_1,\ldots,x_n) = (x_1 + x_2, x_2 + x_3, \ldots, x_{n-1} + x_n)$.

(3) $m = 2$, $n = 3h$, $t = 2h - 1$. Define

$$f(x_1,\ldots,x_{3h}) = (x_1 + \ldots + x_{2h}, x_{h+1} + \ldots + x_{3h}).$$

In fact, all three of these examples are optimal. It is easy to see that $n \ge m + t$, so the first two examples are optimal. The result that $t < \lfloor \frac{2n}{3} \rfloor$ if $m = 2$ is much more difficult; it is proved in [4].

2 Resilient functions and orthogonal arrays

Resilient functions turn out to be equivalent to certain large sets of orthogonal arrays, which we now define. 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 pairs of symbols occurs in exactly λ rows. If $\lambda = 1$, then we write $OA(t, k, v)$.

An orthogonal array is said to be *rowwise simple* if no two rows are identical. Of course, an array with $\lambda = 1$ is rowwise simple. In this paper, we consider only rowwise simple arrays.

A *large set* of orthogonal arrays $OA_\lambda(t, k, v)$ is defined to be a set of v^{k-t}/λ rowwise simple arrays $OA_\lambda(t, k, v)$ such that every possible k−tuple of symbols occurs in exactly one of the OA's in the set. (Equivalently, the union of the OA's forms an $OA(k, k, v)$.)

Here is our main result.

Theorem 2.1 *An (n, m, t)−resilient function is equivalent to a large set of orthogonal arrays $OA_{2^{n-m-t}}(t, n, 2)$.*

Proof. First, suppose $f : \{0,1\}^n \rightarrow \{0,1\}^m$ is an (n, m, t)−resilient function. For any $y \in \{0,1\}^m$, form an array A_y whose rows are the vectors in the inverse image $f^{-1}(y)$. A_y is a $|f^{-1}(y)| \times n$ binary array. It is clear that the 2^m arrays A_y together contain every possible n–tuple as a row, so if each A_y is an $OA_{2^{n-m-t}}(t,n,2)$, then we automatically get a large set.

Let $\{i_1,\ldots,i_t\} \subseteq \{1,\ldots,n\}$ be a t–subset, and let $z_j \in \{0,1\}$ $(1 \leq j \leq t)$. For every $y \in \{0,1\}^m$, let $\lambda(y)$ denote the number of rows in A_y in which z_j occurs in column i_j for $1 \leq j \leq t$. It is easy to see that

$$\sum_{y\in\{0,1\}^m} \lambda(y) = 2^{n-t}.$$

Now

$$p(f(x_1,\ldots,x_n) = (y_1,\ldots,y_m)|x_{i_j} = z_j, 1 \leq j \leq t) = \frac{\lambda(y)}{2^{n-t}}.$$

Since f is t–resilient, we get

$$\frac{\lambda(y)}{2^{n-t}} = \frac{1}{2^m},$$

or $\lambda(y) = 2^{n-m-t}$. Since $\{i_1,\ldots,i_t\}$ and z_j $(1 \leq j \leq t)$ are arbitrary, we have shown that each A_y is an $OA_{2^{n-m-t}}(t,n,2)$, as desired.

Conversely, suppose we start with a large set of $OA_{2^{n-m-t}}(t,n,2)$. There are 2^m arrays in the large set; name them A_y, $y \in \{0,1\}^m$. Then define a function $f : \{0,1\}^n \to \{0,1\}^m$ by the rule

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

It is easy to see that the function f is t–resilient. □

Remark. The fact that the t–resilient function gives a large set of orthogonal arrays was remarked in [4, p. 402].

As an illustration, consider Example (3) in Section 1 with $h = 2$:

$$f(x_1,x_2,x_3,x_4,x_5,x_6) = (x_1 + x_2 + x_3 + x_4, x_3 + x_4 + x_5 + x_6),$$

where addition is modulo 2. This is a $(6,2,3)$–resilient function, and by Theorem 2.1, it is equivalent to a large set of $OA_2(3,6,2)$. There are four OA's in the large set, one of which is obtained from $f^{-1}(0,0)$:

0	0	0	0	0	0		0	1	0	1	0	1
0	0	0	0	1	1		0	1	0	1	1	0
1	1	0	0	0	0		1	0	0	1	0	1
1	1	0	0	1	1		1	0	0	1	1	0
0	0	1	1	0	0		0	1	1	0	0	1
0	0	1	1	1	1		0	1	1	0	1	0
1	1	1	1	0	0		1	0	1	0	0	1
1	1	1	1	1	1		1	0	1	0	1	0

A related result for correlation-immune functions was proved in [3]:

Theorem 2.2 *A correlation-immune function $f : \{0,1\}^n \to \{0,1\}$ of order t is equivalent to an orthogonal array $OA_\lambda(t,n,2)$ for some integer λ.*

Theorem 2.2 can be proved in a similar way as Theorem 2.1 (however, the proof in [3] is very different, making use of a Walsh transform characterization of correlation-immune functions). In fact, we get two orthogonal arrays: an $OA_{\lambda_0}(t,n,2)$ from

$f^{-1}(0)$ and an $OA_{\lambda_1}(t,n,2)$ from $f^{-1}(1)$. For $i = 0,1$, we have $\lambda_i = |f^{-1}(i)|/2^t$, and the union of the two orthogonal arrays is an $OA(k,k,n)$.

In view of Theorem 2.1, any necessary condition for the existence of an orthogonal array $OA_{2^{n-m-t}}(t,n,2)$ is also a necessary condition for the existence of an $(n,m,t)-$resilient function. One classical bound for orthogonal arrays is the Rao bound [9], proved in 1947. We record the Rao bound as the following theorem.

Theorem 2.3 *Suppose there exists an $OA_\lambda(t,k,v)$. Then*

$$\lambda v^t \geq 1 + \sum_{i=1}^{t/2} \binom{k}{i}(v-1)^i$$

if t is even; and

$$\lambda v^t \geq 1 + \sum_{i=1}^{(t-1)/2} \binom{k}{i}(v-1)^i + \binom{k-1}{(t-1)/2}(v-1)^{(t+1)/2}$$

if t is odd.

We obtain the following corollary which gives a necessary condition for existence of a an $(n,m,t)-$resilient function.

Corollary 2.4 *Suppose there exists an $(n,m,t)-$resilient function. Then*

$$m \leq n - \log_2 \left[\sum_{i=0}^{t/2} \binom{k}{i}\right]$$

if t is even; and

$$m \leq n - \log_2 \left[\sum_{i=0}^{(t-1)/2} \binom{k}{i} + \binom{k-1}{(t-1)/2}\right]$$

if t is odd.

Proof. Set $v = 2$ in Theorem 2.3 and apply Theorem 2.1. □

Remark. For t even, the bound of Corollary 2.4 was proved in [4] from first principles. For t odd, our bound is a slight improvement over the bound in [4].

The Bush bound for orthogonal arrays with $\lambda = 1$ [2] also will provide a necessary existence condition for certain resilient functions. This bound is as follows:

Theorem 2.5 *[2] Suppose there exists an $OA(t,k,v)$, where $t > 1$. Then*

$$\begin{array}{lll} k & \leq & v+t-1 \quad \text{if } v \geq t,\ v \text{ even} \\ k & \leq & v+t-2 \quad \text{if } v \geq t \geq 3,\ v \text{ odd} \\ k & \leq & t+1 \qquad\quad \text{if } v \leq t. \end{array}$$

As a corollary, we can obtain the following result that was proved in [1] from first principles:

Corollary 2.6 *[1] There exists an $(n, m, t)-$resilient function with $n = m + t$ if and only if $t = 1$ or $m = 1$.*

Proof. The cases $t = 1$ and $m = 1$ were given earlier in examples. So, suppose $n = m + t$ and $2 \leq t \leq n - 2$. Apply Theorem 2.5 with $v = 2$ to get $m + t \leq t + 1$, or $m \leq 1$, a contradiction. □

3 Resilient functions and error-correcting codes

The most important construction method for resilient functions uses (linear) binary codes. We will be using several standard results from coding theory without proof; see MacWilliams and Sloane [7] for background information on error-correcting codes. An (n, m, d) linear code is an $m-$dimensional subspace C of $(GF(2))^n$ such that any two vectors in C have Hamming distance at least d. Let G be an $m \times n$ matrix whose rows form a basis for C; G is called a *generating matrix* for C. The following construction for resilient functions was given in [1, 4]:

Theorem 3.1 *Let G be a generating matrix for an (n, m, d) linear code C. Define the function $f : (GF(2))^n \rightarrow (GF(2))^m$ by the rule $f(x) = xG^T$. Then f is an $(n, m, d - 1)-$resilient function.*

This result can easily be seen to be true using the orthogonal array characterization. The inverse image $f^{-1}(0, \ldots, 0)$ is in fact the dual code $C^{\perp}$. It is well-known that $C^{\perp}$ is an orthogonal array $OA_{2^{n-m-d+1}}(d - 1, n, 2)$ (see for example [7, p. 139]). In fact, this is obvious since any $d-1$ columns of the generating matrix for $C^{\perp}$ (= the parity check matrix for C) are linearly independent. Now, any other inverse image $f^{-1}(y)$ is an additive coset of $C^{\perp}$, and thus is also an $OA_{2^{n-m-d+1}}(d-1, n, 2)$. Hence we obtain 2^m OA's that form a large set. By Theorem 2.1, f is an $(n, m, d-1)-$resilient function.

As an example, suppose we start with the perfect binary Hamming code [7, p. 25]. This is an $(2^r - 1, 2^r - r - 1, 3)$ code. It gives rise to a $(2^r - 1, 2^r - r + 1, 2)$ resilient function; or equivalently, a large set of orthogonal arrays $OA_{2^{r-2}}(2, 2^r - 1, 2)$. These resilient functions are optimal in view of Corollary 2.4.

As another example, suppose we start with the Reed-Muller code $\mathcal{R}(1, s)$ [7, p. 376]. This is a $(2^s, s+1, 2^{s-1})$ linear code, which yields a $(2^s, s+1, 2^{s-1} - 1)-$resilient function. (Note that a $(2^s, s, 2^{s-1} - 1)-$resilient function is constructed in [4]. This function corresponds to the code obtained from $\mathcal{R}(1, s)$ by deleting the row $1, 1, \ldots, 1$ from the generating matrix. So we get one more output bit than [4], while maintaining the same resiliency.)

Here is an interesting question for future research. It is conceivable that a (rowwise simple) orthogonal array might exist, but a large set (= resilient function) does not. One interesting situation where this might happen concerns the parameters $n = 3h$, $m = 2$, $t = 2h$. It was mentioned earlier that there is no resilient function with these parameters. But the proof of this fact, which is found in [4], does not seem to rule out the existence of an $OA_{2^{h-2}}(2h, 3h, 2)$. So this is a case where an OA might exist even though the large set does not.

In fact, there is no $OA_{2^{h-2}}(2h, 3h, 2)$ if $h = 2$ or $h = 3$, as can be seen by applying the Rao bound. But for $h \geq 4$, it seems that no results are known concerning this class of OA's.

Finally, we mention that Teirlinck has observed in [12] that existence of an orthogonal array $OA(t, k, v)$ (with $\lambda = 1$) implies the existence of a large set of $OA(t, k, v)$. Also, recent results of Friedman [5] show that, for certain other parameter situations, existence of an OA implies the existence of a large set.

Acknowledgements

D. R. Stinson's research is supported by NSF Grant CCR-9121051.

References

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

[2] K. A. Bush. Orthogonal arrays of index unity. *Ann. Math. Stat.* **23** (1952), 426-434.

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

[4] B. Chor, O. Goldreich, J. Hastad, J. Friedman, S. Rudich and R. Smolensky. The bit extraction problem or t−resilient functions. *Proc. 26th IEEE Symp. on Foundations of Computer Science* (1985), 396-407.

[5] J. Friedman. On the bit extraction problem. *Proc. 33rd IEEE Symp. on Foundations of Computer Science* (1992), 314-319.

[6] X. Guo-zhen and J. L. Massey. A spectral characterization of correlation-immune functions. *IEEE Trans. Inform. Theory* **34** (1988), 569-571.

[7] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-correcting Codes*, North-Holland, 1977.

[8] C. R. Rao. Hypercubes of strength "*d*" leading to confounded designs in factorial experiments. *Bull. Calcutta Math. Soc.* **38** (1946), 67-78.

[9] C. R. Rao. Factorial experiments derivable from combinatorial arrangements of arrays. *J. Royal Stat. Soc.* **9** (1947), 128-139.

[10] R. Rueppel. *Analysis and Design of Stream Ciphers*, Springer Verlag, Berlin, 1986.

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

[12] L. Teirlinck. Large sets of disjoint designs and related structures. In *Contemporary Design Theory – A Collection of Surveys*, John Wiley & Sons, New York, 1992, 561-592.