Covering arrays of strength three*

M.A. Chateauneuf
Department of Mathematical Sciences,
Michigan Technological University,
Houghton, Michigan 49931-1295.

Charles J. Colbourn
Department of Computer Science,
University of Vermont,
Burlington, Vermont 05405.

D. L. Kreher
Department of Mathematical Sciences,
Michigan Technological University,
Houghton, Michigan 49931-1295.

Abstract

A *covering array* of *size* N, *degree* k, *order* v and *strength* t is a $k \times N$ array with entries from a set of v symbols such that in any $t \times N$ subarray every $t \times 1$ column occurs at least once. Covering arrays have been studied for their applications to drug screening and software testing. We present explicit constructions and give constructive upper bounds for the size of a covering array of strength three.

Keywords: covering array, orthogonal array, group action, perfect hash family.

1 Introduction

A *covering array* of *size* N, *degree* k, *order* v and *strength* t is a $k \times N$ array with entries from a set of v symbols such that in any $t \times N$ subarray every $t \times 1$ column occurs at least once. We denote such an array by $\mathsf{CA}(N; t, k, v)$. The *covering array number* $\mathsf{CAN}(t, k, v)$ is the fewest columns N in a $\mathsf{CA}(N; t, k, v)$. An obvious lower bound is:

$$v^t \leq \mathsf{CAN}(t, k, v). \tag{1}$$

*Designs, Codes and Cryptography, 16 235–242 (1999).

Suppose A is a *covering array* of type $\mathsf{CA}(N; t, k, v)$ and let i be any row and x any symbol. Then the $k-1 \times N'$ subarray obtained by deleting row i from A and keeping only those columns of A that had symbol x in row i is a $\mathsf{CA}(N'; t-1, k-1, v)$, where N' is the number of occurrences of x in row i. Hence

$$\mathsf{CAN}(t-1, k-1, v) \leq \frac{1}{v}\mathsf{CAN}(t, k, v). \tag{2}$$

We present new explicit constructions for covering arrays of strength three and obtain a general upper bound on the covering number. Some of the techniques used are similar to those used for constructing orthogonal arrays [5, 9]. Covering arrays have a number of applications in experimental design [11], for example in software testing [2, 3] and in drug screening [13]. In these applications, it is imperative that the interaction of all combinations of t parameters be tested, and hence that all such selections are covered by columns of the array. This is complementary to standard error-correcting codes in which columns are not to cover a selection more than once [11].

2 Some constructions

Let Ω be the set of v symbols on which we are to construct a $\mathsf{CA}(N; 3, k, v)$. Let G be a group acting on the set Ω. If $g \in G$ and M is a $k \times \ell$ matrix with entries in Ω, then M^g is the $k \times \ell$ matrix whose $[i, j]$ entry is $M[i, j]^g$, the image of $M[i, j]$ under g. The matrix obtained by developing M by G is the $k \times \ell|G|$ matrix

$$M^G = [M^g : g \in G].$$

Let $C = C(k, \Omega)$ be the $k \times |\Omega|$ matrix that has a constant column with each entry equal to x, for each $x \in \Omega$. When the k rows are indexed by a set X, the notation $C(X, \Omega)$ is also used. The goal is to choose the matrix M and group G so that $[M^G, C]$ or just $[M^G]$ is a $\mathsf{CA}(N; 3, k, v)$. For example, when $G = \mathsf{Sym}\{0, 1, 2\}$ (the symmetric group on $\{0, 1, 2\}$), and

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 2 & 1 & 0 \end{bmatrix},$$

$[M^G, C]$ is a $\mathsf{CAN}(3, 4, 3)$.

Theorem 2.1 *Let $v > 2$ be a positive integer, and let $q \geq v - 1$ be a prime power. Then there is a $\mathsf{CA}((2v-1)(q^3-q)+v; 3, 2v, v)$.*

Proof: First we construct a $\mathsf{CA}(N; 3, 2v, q+1)$, with

$$N = (2v-1)(q^3-q) + q + 1.$$

Since q is a prime power, the group $G = PGL(q)$ is sharply 3-transitive on the projective line $\Omega = \mathsf{GF}(q) \cup \{\infty\}$. Under this group there are precisely five orbits of 3-tuples. These five orbits are determined by the pattern of the entries in their 3-tuples:

1. $\{[a,a,a]^T : a \in \Omega\}$
2. $\{[a,a,b]^T : a,b \in \Omega, a \neq b\}$
3. $\{[a,b,a]^T : a,b \in \Omega, a \neq b\}$
4. $\{[b,a,a]^T : a,b \in \Omega, a \neq b\}$
5. $\{[a,b,c]^T : a,b,c \in \Omega, a \neq b \neq c \neq a\}$

Let x_1, x_2, x_3 be any three rows of $[M^G, C(X,\Omega)]$. The 3-tuples with all equal entries occur on rows x_1, x_2, x_3 of $[M^G, C(X,\Omega)]$ since they occur in $C(X,\Omega)$. Thus to construct a $\mathsf{CA}(N;3,2v,q+1)$ using this group we need only find a matrix M such that for each of the orbits 2-5 and each set of 3 rows there is a column of M that contains an orbit representative for the orbit on the chosen rows.

Let X be a $2v$-element set of vertices and let

$$\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_{2v-1}$$

be a one-factorization of the complete graph $\mathcal{G}$ on X. Let $\ell : E(\mathcal{G}) \to \Omega$be an arbitrary function for which $\ell(e) \neq \ell(e')$ whenever $e, e' \in \mathcal{F}_j$ for some j. Such a function exists since $|\Omega| = q+1 > v$. Define the $2v \times (2v-1)$ matrix M with entries in Ω by $M[x,j] = \ell(e)$, where e is the edge of $\mathcal{F}_j$ that is incident to $x \in X$.

It must be shown that each of the orbits 2, 3, 4, and 5 has a representative on rows x_1, x_2, x_3. The edge $e = \{x_1, x_2\}$ is an edge of some one-factor $\mathcal{F}_j$ and x_3 is incident to some other edge e' of $\mathcal{F}_j$. Thus $\ell(e) = a$ and $\ell(e') = b$ for some $a \neq b$ in Ω. Consequently

$$[M[x_1,j], M[x_2,j], M[x_3,j]] = [a,a,b]$$

and so orbit (2) is represented. Similarly, orbits (3) and (4) are represented. There are $2v-4$ one-factors that do not contain any of the edges $\{x_1,x_2\}, \{x_1,x_3\}$, or $\{x_2,x_3\}$. Thus, since $2v \geq 5$, there is a one-factor $\mathcal{F}_{j'}$ in which x_1, x_2, and x_3 are each incident to different edges. Consequently, column j' of M has distinct entries on rows x_1, x_2, and x_3 and thus orbit (5) is represented. Therefore $[M^G, C(X,\Omega)]$ is a $\mathsf{CA}(N;3,2v,q+1)$, with $N = (2v-1)(q^3-q)+q+1$. To obtain a $\mathsf{CA}(N';3,2v,v)$ with $N' = (2v-1)(q^3-q)+v$ replace $q+1-v$ of the non-zero symbols with 0 and delete the $q+1-v$ extra columns of 0's in $C(X,\Omega)$. □

When $q = 2$, $PGL(2,q)$ is isomorphic to $\mathsf{Sym}\{0,1,2\}$. Figure 1 depicts the $\mathsf{CA}(33;3,6,3)$ array that is constructed using this group by Theorem 2.1.

Theorem 2.2 *The covering number* $\mathsf{CAN}(3,6,3) = 33$.

Proof: Östergård has shown that $\mathsf{CAN}(2,5,3) = 11$ (see [11]). Thus by inequality (2), $\mathsf{CAN}(3,6,3) \geq 33$. But by Theorem 2.1, $\mathsf{CAN}(3,6,3) \leq 33$. □

Example 2.3 illustrates a situation in which a good array is obtained without using 3-transitive group.

0 1 2 2 1	1 2 0 0 2	2 0 1 1 0	0 2 1 1 2	1 0 2 2 0	2 1 0 0 1	0 1 2
1 2 2 1 0	2 0 0 2 1	0 1 1 0 2	2 1 1 2 0	0 2 2 0 1	1 0 0 1 2	0 1 2
2 2 1 0 1	0 0 2 1 2	1 1 0 2 0	1 1 2 0 2	2 2 0 1 0	0 0 1 2 1	0 1 2
2 1 0 1 2	0 2 1 2 0	1 0 2 0 1	1 2 0 2 1	2 0 1 0 2	0 1 2 1 0	0 1 2
1 0 1 2 2	2 1 2 0 0	0 2 0 1 1	2 0 2 1 1	0 1 0 2 2	1 2 1 0 0	0 1 2
0 0 0 0 0	1 1 1 1 1	2 2 2 2 2	0 0 0 0 0	1 1 1 1 1	2 2 2 2 2	0 1 2

Figure 1: A minimal $\mathsf{CA}(33; 3, 6, 3)$ covering array

Example 2.3 *A* $\mathsf{CA}(88; 3, 8, 4)$ *covering array.*

Construction: Let $X = \mathsf{GF}(8) = \mathbb{Z}_2[x]/(x^3+x+1)$ be the points of the complete graph K_8. Observe that x is a primitive root of X. Let $\mathcal{F}$ be the cosets of $H = \{0, 1\}$ as a subgroup of the additive group of $\mathsf{GF}(8)$. Then

$$\mathcal{F} = \{\{0, 1\}, \{x, x^3\}, \{x^2, x^6\}, \{x^4, x^5\}\}$$

is a one factor and the partition

$$\{x\mathcal{F}, x^2\mathcal{F}, x^3\mathcal{F}, x^4\mathcal{F}, x^5\mathcal{F}, x^6\mathcal{F}\}$$

is a one-factorization of K_8. Let $\ell : E(K_8) \to \Omega$ be any function such that $\ell(e) \neq \ell(e')$ whenever $e, e' \in x^j\mathcal{F}$ for some j and define the 8×7 matrix M with entries in Ω by $M[x, j] = \ell(e)$ where e is the edge of $x^j\mathcal{F}$ that is incident to $x \in X$. One such matrix is given in Figure 2.

Let $G = \mathsf{Alt}\Omega$ be the alternating group of permutations of Ω. Then $[M^G, C(X, \Omega)]$ is the desired $\mathsf{CA}(88; 3, 8, 4)$ covering array. To see this, consider any three rows x_1,x_2,x_3. It must be shown that, for each choice of a, b, c with $a \neq b \neq c \neq a$, each of the patterns $[a, a, a]^T$, $[a, a, b]^T$, $[a, b, a]^T$, $[b, a, a]^T$, and $[a, b, c]^T$ occurs on these rows. Patterns with three equal entries occur in $C = C(X, \Omega)$. Patterns with two equal entries and one different appear in $[M^G, C(X, \Omega)]$ on rows x_1, x_2, x_3 since G is 2-transitive on Ω and every pair occurs as an edge. This leaves only patterns with three distinct entries. There are exactly two orbits of such patterns under G since G is the alternating group on four symbols. Thus it suffices that in M there are two triples $[x, y, z]^T$ and $[x', y', z']^T$ on rows i, j, k for which the unique permutation in the symmetric group that sends one on to the other is an odd permutation. This is easily checked. □

Corollary 2.4 *The covering number* $\mathsf{CAN}(3, 8, 4) \leq 88$ *and* $\mathsf{CAN}(2, 7, 4) \leq 22$.

Proof: The construction in Example 2.3 and the inequality (2) establish this result. □

Stevens and Mendelsohn [12] give recursive techniques that can be used to establish the bound $\mathsf{CAN}(2, 7, 4) \leq 25$, upon which this improves. It is somewhat surprising that the construction for $t = 3$ is sufficiently strong to improve upon this bound for the derived arrays when $t = 2$.

0	0	0	0	0	0	0
0	2	3	3	1	2	1
1	0	2	3	3	1	2
2	1	0	2	3	3	1
1	2	1	0	2	3	3
3	1	2	1	0	2	3
3	3	1	2	1	0	2
2	2	2	1	2	1	0

Figure 2: A matrix for the one-factorization in Construction 2.3

3 Constructive upper bounds

A $(k, n; \lambda)$-*difference matrix* is a $n \times k\lambda$ matrix D, with entries in $\mathbb{Z}_k$, in which the vector difference of any pair of rows contains every member of $\mathbb{Z}_k$ exactly λ times. For example, if $\gcd((n-1)!, k) = 1$, then the $n \times k$ matrix D defined by $D[i, j] = ij \bmod k$ is a $(k, n; 1)$-difference matrix.

Lemma 3.1 (Atici et. al. [1]) *Suppose $T_1, T_2, \ldots, T_r \subseteq \mathbb{Z}_k$, where $T_i \neq \emptyset$, $1 \leq i \leq r$ and*

$$\sum_{j=1}^{r} |T_j| = t.$$

Suppose also that D is a $(k, \binom{t}{2} + 1; 1)$-difference matrix. Then there is an integer x such that the r sets

$$T'_j = \{t + D[x, j] : t \in T_j\}$$

$1 \leq j \leq r$ are all disjoint.

We use Lemma 3.1 to establish a slight generalization of Theorem 4.1 in [1]. Let Ω be a v-element set and let $\mathcal{P}$ be a first-order predicate concerning a collection $\mathcal{C}$ of t-tuples with entries in Ω that is invariant under coordinate permutation. For example the predicate $\mathcal{P}$ could be one of

1. The t-tuple $[a, a, ..., a]$ is in $\mathcal{C}$, where $a \in \Omega$ is a fixed symbol.
2. The set of all t-tuples with exactly two distinct entries are in $\mathcal{C}$.
3. There is a t-tuple in $\mathcal{C}$ with distinct entries.
4. All of the t-tuples with entries in Ω occur in $\mathcal{C}$.

A $k \times N$ array A with entries in Ω is called a $(N; t, k, v)$-$\mathcal{P}$ array if each collection $\mathcal{C}$ of t-tuples found among the columns of any $t \times N$ subarray satisfies the predicate $\mathcal{P}$. For example if $\mathcal{P}$ is predicate 3 then the array A is a perfect hash family [1] or a t-separating family [10]. If $\mathcal{P}$ is predicate 4, then A is a covering array.

Theorem 3.2 *If* $\gcd(\binom{t}{2}!, k) = 1$ *and a* $(N; t, k, v)$*-*$\mathcal{P}$ *array exists, then there is a* $((\binom{t}{2} + 1)N, t, k^2, v)$*-*$\mathcal{P}$ *array.*

Proof: (This proof parallels one in [1].) Let A be a $(N; t, k, v)$-$\mathcal{P}$ array with rows labeled by $\mathbb{Z}_k$ and columns by $\{1, 2, \ldots, N\}$. Let D be a $(k, \binom{t}{2}+1; 1)$-difference matrix and for $x \in \mathbb{Z}_k$, let A_x denote the array

$$A_x[i, j] = A[i + x, j], \qquad i \in \mathbb{Z}_k, \quad j = 1, 2, ..., N$$

Consider the array B defined by

$$B = \begin{array}{|c|c|c|c|} \hline B_{0,0} & B_{0,1} & \cdots & B_{0,\binom{t}{2}} \\ \hline B_{1,0} & B_{1,1} & \cdots & B_{1,\binom{t}{2}} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline B_{k-1,0} & B_{k-1,1} & \cdots & B_{k-1,\binom{t}{2}} \\ \hline \end{array}$$

where $B_{i,j} = A_{D[i,j]}$ and the rows of B are indexed by $\mathbb{Z}_k \times \mathbb{Z}_k$ so that the rows of

$$\begin{array}{|c|c|c|c|} \hline B_{i,0} & B_{i,1} & \cdots & B_{i,\binom{t}{2}} \\ \hline \end{array}$$

are indexed by $\{i\} \times \mathbb{Z}_k$. To see that B is the desired $((\binom{t}{2} + 1)N; t, k, v)$-$\mathcal{P}$ array consider a set of t rows $\mathcal{R}$ of B. Set $T_i = \{x \in \mathbb{Z}_k : (i, x) \in \mathcal{R}\}$. Let $T_{i_1}, T_{i_2}, \ldots, T_{i_r}$ denote the T_i's that are nonempty. Then by Lemma 3.1 there is an integer x such that the r sets $T'_{i_j} = \{t + D[x, j] : t \in T_{i_j}\}$ for $1 \le j \le r$ are all disjoint. Thus they comprise t distinct rows of the $k \times (\binom{t}{2} + 1)N$ subarray

$$\begin{array}{|c|c|c|c|} \hline B_{x,0} & B_{x,1} & \cdots & B_{x,\binom{t}{2}} \\ \hline \end{array}$$

Each component $B_{x,j}$ is a copy of A with the rows permuted. Thus the t rows satisfy predicate $\mathcal{P}$. Therefore B is a $((\binom{t}{2} + 1)N, t, k^2, v)$-$\mathcal{P}$ array. □

This theorem can be iterated:

Theorem 3.3 *If* $\gcd(\binom{t}{2}!, k) = 1$ *and a* $(N; t, k, v)$*-*$\mathcal{P}$ *array exists, then there is a* $(N(\binom{t}{2} + 1)^j, t, k^{2^j}, v)$*-*$\mathcal{P}$ *array.*

Applying this to covering arrays we obtain:

Theorem 3.4 *Let $v > 2$ be a positive integer, and let $q \geq v - 1$ be a prime power. Then, for all j, there is a $\mathsf{CA}(((2v-1)(q^3-q)+v)4^j; 3, (2v-1)^{2^j}, v)$ if $v \equiv 0, 1 \bmod 3$, and a $\mathsf{CA}(((2v-1)(q^3-q)+v)4^j; 3, (2v-3)^{2^j}, v)$ if $v \equiv 2 \bmod 3$.*

Proof: There is, by Theorem 2.1, a $\mathsf{CA}((2v-1)(q^3-q)+v; 3, 2v, v)$. Deleting rows leaves a covering array. If $v \equiv 0, 1 \bmod 3$, then $\gcd(6, 2v-1) = 1$ and if $v \equiv 2 \bmod 3$, then $\gcd(6, 2v-3) = 1$. Apply Theorem 3.3 with $t = 3$ and $k = 2v-1$ when $v \equiv 0, 1 \bmod 3$ or $k = 2v-3$ when $v \equiv 2 \bmod 3$. □

With the parameters in Theorem 3.4, we have

$$\mathsf{CAN}(3, k, v) \leq \begin{cases} \frac{(2v-1)(q^3-q)+v}{(\log(2v-1))^2}(\log k)^2 \text{ if } v \equiv 0, 1 \bmod 3 \\ \frac{(2v-1)(q^3-q)+v}{(\log(2v-3))^2}(\log k)^2 \text{ if } v \equiv 2 \bmod 3 \end{cases}$$

where $q \geq v - 1$ is a prime power. In particular when $v = 3$ we have

$$\mathsf{CAN}(3, k, 3) \leq \frac{33}{(\log 5)^2}(\log k)^2 \approx 6.1209(\log k)^2,$$

and when $v = 4$ we have

$$\mathsf{CAN}(3, k, 4) \leq \frac{172}{(\log 7)^2}(\log k)^2 \approx 21.8240(\log k)^2.$$

For $v = 4$ we can improve this using the $\mathsf{CA}(88; 3, 8, 4)$ from Example 2.3:

$$\mathsf{CAN}(3, k, 4) \leq \frac{88}{(\log 7)^2}(\log k)^2 \approx 11.1658(\log k)^2$$

The bound can be improved asymptotically to be on the order of $v^2 \log k$ using probabilistic techniques [7]. However, the technique developed here is entirely constructive, and hence can lead to useful small covering arrays for use in experimental design.

Acknowledgments

Thanks to Brett Stevens and Jeff Dinitz for helpful discussions, and to Doug Stinson for helpful comments on the paper. The authors' research is supported as follows: ARO grant DAAG55-98-1-0272 (Colbourn) and NSA grant MDA904-97-1-0072 (Kreher).

References

[1] M. Atici, S. S. Magliveras, D. R. Stinson, and W.-D. Wei, Some recursive constructions for perfect hash functions, *J. Combinat. Designs* **4** (1996), 353-363.

[2] D. M. Cohen, S. R. Dalal, M. L. Fredman, and G. C. Patton, The AETG system: an approach to testing software based on combinatorial design, *IEEE Trans. Software Engineering* **23** (1997), 437-444.

[3] D. M. Cohen, S. R. Dalal, J. Parelius, and G. C. Patton, The combinatorial design approach to automatic test generation, *IEEE Software* **13** (1996), 83-88.

[4] C. J. Colbourn and J. H. Dinitz (editors), *The CRC Handbook of Combinatorial Designs*, CRC Press, Boca Raton, 1996.

[5] C. J. Colbourn and D. L. Kreher, Concerning difference matrices, *Designs, Codes and Cryptography* **9** (1996), 61-70.

[6] L. Gargano, J. Körner, and U. Vaccaro, Sperner capacities, *Graphs and Combinatorics* **9** (1993), 31-46.

[7] A. P. Godbole, D. E. Skipper, and R. A. Sunley, t-Covering arrays: upper bounds and Poisson approximations, *Combinatorics, Probability and Computing* **5** (1996), 105-118.

[8] J. Körner and M. Lucertini, Compressing inconsistent data, *IEEE Trans. Information Theory* **40** (1994), 706-715.

[9] D. L. Kreher, Orthogonal arrays of strength 3, *J. Combinat. Designs* **4** (1996), 67-69.

[10] S. Poljak, A. Pultr, and V. Rödl, On qualitatively independent partitions and related problems, *Discrete Applied Math.* **6** (1983), 193-205.

[11] N. J. A. Sloane, Covering arrays and intersecting codes, *J. Combinat. Designs* **1** (1993), 51-63.

[12] B. Stevens and E. Mendelsohn, New recursive methods for transversal covers, *J. Combinat. Designs*, to appear.

[13] A.-I. Tong, Y.-G. Wu, and L.-D. Li, Room-temperature phosphorimetry studies of some addictive drugs following dansyl chloride labelling, *Talanta* 43 (1996), 1429-1436.