UNION DISTINCT FAMILIES OF SETS, WITH AN APPLICATION TO CRYPTOGRAPHY

Mausumi Bose
Indian Statistical Institute,
203 B.T. Road, Kolkata 700 108, India
E-mail: mausumi.bose@gmail.com
and
Rahul Mukerjee
Indian Institute of Management Calcutta
Joka, Diamond Harbour Road, Kolkata 700 104, India
E-mail: rmuk0902@gmail.com

Abstract: A family of sets is called K-union distinct if all unions involving K or fewer members thereof are distinct. If a family of sets is K-cover-free then it is K-union distinct. In this paper, we recognize that this is only a sufficient condition and, from this perspective, consider partially cover-free families of sets with a view to constructing union distinct families. The role of orthogonal arrays and related combinatorial structures is explored in this context. The results are applied to find efficient anti-collusion digital fingerprinting codes.
Key words: Anti collusion code, constant weight code, cover-free family, orthogonal array, partially cover-free family, resilience.

1. Introduction

Cover-free families of sets, introduced by Kautz and Singleton [6] have received significant attention in the literature, with application to such diverse fields as group testing, cryptography and communications. Let Ω be a universal set of v elements and $\mathcal{H}$ be a family consisting of n subsets of Ω. Then $\mathcal{H}$ is called a *K-cover-free family*, or K-CFF(v, n), if no union of K or fewer members of $\mathcal{H}$ includes as a subset, i.e., covers, any member of $\mathcal{H}$ other than those involved in the union. It is readily seen that if $\mathcal{H}$ is a K-CFF(v, n), then $\mathcal{H}$ is also a *K-union distinct family*, or K-UDF(v, n), in the sense that all unions involving K or fewer members of $\mathcal{H}$ are distinct. Indeed, it is this feature that accounts for certain applications of cover-free families, notably in the construction of anti-collusion digital fingerprinting codes; see [11]. Hence we focus on union distinct families of sets in the present article. While such families are of theoretical interest on their own, they will be seen to be attractive also from practical considerations.

Although a K-cover-free family of sets is K-union distinct, the converse is not true. The following example shows that a family of sets can be K-union distinct even without being K-cover-free.

Example 1. Let $\Omega = \{1, 2, 3\} \times \{0, 1, 2\}$, where $\times$ denotes Cartesian product of sets. Consider a family $\mathcal{H}$ of $n = 12$ subsets of Ω as given by

$\{10, 20, 30\}$, $\{10, 21, 32\}$, $\{10, 22, 31\}$, $\{11, 20, 32\}$, $\{11, 21, 31\}$, $\{11, 22, 30\}$, $\{12, 20, 31\}$, $\{12, 21, 30\}$, $\{12, 22, 32\}$, $\{10, 21, 31\}$, $\{11, 22, 32\}$, $\{12, 20, 30\}$.

Then $\mathcal{H}$ is not 2-cover-free – e.g., the union of {10, 20, 30} and {11, 21, 31} covers the set {10, 21, 31}. Nevertheless, one can directly verify that $\mathcal{H}$ is 2-union distinct. It may also be noted that $\mathcal{H}$ is only *partially* 2-cover-free in the sense that the subfamily, consisting of the first nine sets, is indeed 2-cover-free.

Clearly, if $\mathcal{H}$ is K-union distinct then so is any subfamily of $\mathcal{H}$. Given K and v, therefore, union distinct families with larger n are appealing. Example 1 holds out the promise of the existence of union distinct families which are only partially cover-free but have larger n, for given K and v. Moreover, as we will see, partially cover-free families can also be used to obtain (completely) cover-free and hence union distinct families with larger n. In Section 2, we derive results in this direction on the construction of union distinct families. For this purpose, the notion of union distinct codes is introduced and the role of orthogonal arrays and related combinatorial structures is explored. Finally, in Section 3, the results are applied to find efficient anti-collusion digital fingerprinting codes and illustrative examples are given. The longer proofs appear in the appendix.

In what follows, given a universal set Ω of v elements and a family $\mathcal{H}$ consisting of n subsets of Ω, the *incidence matrix* of $\mathcal{H}$ is defined as a $v \times n$ matrix with (i, j)th element 1 if the ith element of Ω belongs to the jth member of $\mathcal{H}$, and 0 otherwise.

2. Construction of union distinct families

2.1 Union distinct codes

Definition 1. *Let $C = \{(c_{j1},..., c_{jm}) : 1 \le j \le M\}$ be a code consisting of M code vectors each of length m and defined over an alphabet of size s. Then C is called K-union distinct (K-UD) if no two distinct subsets J_1 and J_2 of $\{1,..., M\}$ satisfy $\{c_{ji} : j \in J_1\} = \{c_{ji} : j \in J_2\}$ for every i $(1 \le i \le m)$, when each of J_1, J_2 has K or fewer elements.*

It is not hard to see that if the minimum (Hamming) distance d of C satisfies

$$K(m - d) < m, \tag{1}$$

then C is K-UD. This happens because, with J_1 and J_2 as in Definition 1, there exists a j^* which belongs to one of them, say J_2, but not the other, say J_1. Then for each $j \in J_1$, one has $c_{ji} = c_{j^*i}$ for at most $m - d$ choices of i. Thus $c_{j^*i} \in \{c_{ji} : j \in J_1\}$ for at most $K(m - d)$ choices of i, as J_1 has at most K elements. So, if (1) holds then for some i, one gets $c_{j^*i} \notin \{c_{ji} : j \in J_1\}$ and hence $\{c_{ji} : j \in J_1\} \neq \{c_{ji} : j \in J_2\}$, that is, C is K-UD. The next example, in the spirit of Example 1, shows that C can be K-UD even when (1) does not hold.

Example 2. Let C be a code over an alphabet of size $s = 3$ and consisting of 12 code vectors, each of length $m = 3$, as shown below:

(0, 0, 0), (0, 1, 2), (0, 2, 1), (1, 0, 2), (1, 1, 1), (1, 2, 0),
(2, 0, 1), (2, 1, 0), (2, 2, 2), (0, 1, 1), (1, 2, 2), (2, 0, 0).

The minimum distance of C is $d = 1$, which does not meet (1) for $K = 2$. Still, it may be directly checked that C is 2-UD. The 12 code vectors here correspond to the 12 sets of the family in Example 1, a point which will be clarified below when we continue with this example.

2.2 Union distinct families from union distinct codes

Consider a K-UD code $C = \{(c_{j1},..., c_{jm}) : 1 \le j \le M\}$ over an alphabet $\{\alpha_0, \alpha_1,..., \alpha_{s-1}\}$ of size s. For $1 \le i \le m$ and $1 \le j \le M$, write $\tilde{c}_{ji} = l$ if $c_{ji} = \alpha_l$. Let $\Omega = \{1, 2,..., m\} \times \{0, 1,..., q-1\}$ be a set of mq elements, $\mathcal{F} = \{F(0), F(1),..., F(s-1)\}$ be a family of subsets of $\{0, 1, ..., q-1\}$, and $\mathcal{H}(0)$ be a family consisting of the M subsets

$$E_j = E_{j1} \cup ... \cup E_{jm}, \quad 1 \le j \le M \qquad (2)$$

of Ω, where, for each j,

$$E_{ji} = \{i\} \times F(\tilde{c}_{ji}), \qquad 1 \le i \le m. \qquad (3)$$

Example 2 (continued). We continue with the 2-UD code C in Example 2 to illustrate the ideas underlying (2) and (3). Here $M = 12$ and $m = s = 3$. Let $q = 3$. Then $\Omega = \{1, 2, 3\} \times \{0, 1, 2\} = \{10, 11, 12, 20, 21, 22, 30, 31, 32\}$. The code C considered here is over the alphabet $\{0, 1, 2\}$, i.e., $\alpha_l = l$ $(l = 0, 1, 2)$, so that $\tilde{c}_{ji} = c_{ji}$ for every j and i. Hence taking $F(0) = \{0\}$, $F(1) = \{1\}$, $F(2) = \{2\}$, by (3), we get $E_{ji} = \{i\} \times \{c_{ji}\}$, i.e., E_{ji} is singleton for each j and i, with unique member ic_{ji}. As $M = 12$ and $m = 3$, it now follows from (2) that the family $\mathcal{H}(0)$ consists of the 12 subsets $E_j = \{1c_{j1}, 2c_{j2}, 3c_{j3}\}$ of Ω. For instance, $E_1 = \{10, 20, 30\}$ and $E_2 = \{10, 21, 32\}$ as the first two code vectors in C are (0, 0, 0) and (0, 1, 2). In this manner, one can check that $\mathcal{H}(0)$ is precisely the same as the 2-union distinct family shown in Example 1.

The last example illustrates the construction of a 2-union distinct family via (2) and (3), starting from a 2-UD code and using a smaller family $\mathcal{F} = \{F(0), F(1), F(2)\}$ which is also 2-union distinct. Theorem 1 below puts together these ideas in the form of a general result. We note at this stage that, as (2) and (3) suggest, $\mathcal{H}(0)$ could equivalently be introduced via concatenation of codes but the present description will be more convenient for our proofs.

Theorem 1. *Suppose $\mathcal{F}$ is K-union distinct, that is, a K-UDF(q, s). Then the family $\mathcal{H}(0)$, obtained from the K-UD code C via (2) and (3), is a K-UDF(v, n), where $v = mq$ and $n = M$.*

Proof: If $\mathcal{H}(0)$ is not K-union distinct, then by (2) and (3), there exist two distinct subsets J_1 and J_2 of $\{1,..., M\}$, such that each of J_1 and J_2 has K or fewer elements and, for every i $(1 \le i \le m)$, the union of the sets $F(\tilde{c}_{ji})$, $j \in J_1$, equals the union of the sets $F(\tilde{c}_{ji})$, $j \in J_2$. Since $\mathcal{F}$ is K-union distinct, it fol-

lows that $\{\tilde{c}_{ji} : j \in J_1\} = \{\tilde{c}_{ji} : j \in J_2\}$ and hence $\{c_{ji} : j \in J_1\} = \{c_{ji} : j \in J_2\}$, for $1 \le i \le m$. This is impossible as C is a K-UD code. □

Theorem 1 does not require the K-UD code C to satisfy (1). This is demonstrated by Example 2. Since being K-UD is less stringent than meeting (1), one can hope that this would allow the use of C with larger M and hence yield a union distinct family $\mathcal{H}(0)$ with larger n (= M) via Theorem 1. A construction for such C will be presented in Theorem 3 and the resulting $\mathcal{H}(0)$ will turn out to be only partially cover-free. However, before doing so, we present Theorem 2 below which shows how partially cover-free families can as well be used to obtain (completely) cover-free and hence union distinct families with larger n, even when C satisfies (1). The proof of Theorem 2 appears in the appendix. We continue with the notation for Theorem 1. In addition, we consider a family of subsets $\mathcal{G} = \{G(1),\ldots, G(u)\}$ of $\{0, 1, \ldots, q-1\}$, and for $1 \le i \le m$, define

$$\mathcal{H}(i) = \{\{i\} \times G(1), \ldots, \{i\} \times G(u)\} \qquad (4)$$

as a family consisting of u subsets of Ω.

Theorem 2. *Suppose* (i) $K < m$, (ii) *the minimum distance* d *of* C *satisfies* $K(m-d) < m$, (iii) $\mathcal{F}$ *is K-cover-free, that is, a* K-CFF(q, s), *and* (iv) *no union of* K *or fewer members of* $\mathcal{F} \cup \mathcal{G}$ *covers any member of* $\mathcal{G}$ *other than those involved in the union. Then the family* $\mathcal{H} = \mathcal{H}(0) \cup \mathcal{H}(1) \ldots \cup \mathcal{H}(m)$ *is a* K-CFF(v, n), *and hence a* K-UDF(v, n), *where* $v = mq$ *and* $n = M + mu$.

Remark 1. In Theorem 2, we do not need $\mathcal{F} \cup \mathcal{G}$ to be K-cover-free. While it has a K-cover-free subfamily $\mathcal{F}$, a union of K or fewer members of $\mathcal{F} \cup \mathcal{G}$ can potentially cover a member of $\mathcal{F}$ not involved in the union, and still the theorem remains valid. Thus the family $\mathcal{F} \cup \mathcal{G}$, which is only partially cover-free, leads to a (completely) cover-free and hence union distinct family $\mathcal{H}$. This entails gains in the sense that $\mathcal{H}$ contains more sets than $\mathcal{H}(0)$ which one would have obtained using $\mathcal{F}$ alone. Incidentally, the idea in Theorem 2 of augmenting $\mathcal{H}(1),\ldots, \mathcal{H}(m)$ to $\mathcal{H}(0)$, via the use of $\mathcal{G}$ in addition to $\mathcal{F}$, is reminiscent of an adding-column technique in the construction of orthogonal arrays (see [12]), and this is new in the present context.

Example 3 below illustrates an application of Theorem 2. More examples appear in Section 3. For ease in presenting these examples, we first indicate how orthogonal arrays of index unity and binary codes of constant weight can help in finding $\mathcal{F}$, $\mathcal{G}$ and C, as stipulated in Theorem 2. An orthogonal array OA(s^t, m, s, t) of index unity is an $s^t \times m$ array, with entries from a set of s symbols, such that each ordered t-tuple of symbols occurs exactly once as a row in every $s^t \times t$ subarray.

Proposition 1. Let the rows of an OA(s^t, m, s, t) be taken as the code vectors of C. Then conditions (i) and (ii) of Theorem 1 are met if $t > 1$ and $K(t-1) < m$.

Proof. Obviously here $K < m$, i.e., condition (i) of Theorem 2 is satisfied. Furthermore, following [8], p. 329, the code C obtained as above is maximum distance separable and hence meets the Singleton bound with equality, so that $d = m - t + 1$. Hence $K(m-d) = K(t-1) < m$, and condition (ii) of Theorem 2 is also satisfied. □

We next discuss how one can use binary codes of constant weight to obtain the families $\mathcal{F}$ and $\mathcal{G}$ meeting conditions (iii) and (iv) of Theorem 2. Let $B(q, N, d, w)$ denote a binary code with minimum distance d and N code vectors each of length q and weight w. Consider two such binary codes $B_1 = B(q, s, d_1, w_1)$ and $B_2 = B(q, u, d_2, w_2)$. Write the code vectors of B_1 and B_2 as columns to form $q \times s$ and $q \times u$ matrices and take these as the incidence matrices of $\mathcal{F}$ and $\mathcal{G}$ respectively. If the rows of these matrices are indexed as 0, 1, ..., q–1, then $\mathcal{F}$ and $\mathcal{G}$ contain, respectively, s and u members, each member being a subset of $\{0, 1, \ldots, q-1\}$.

Proposition 2. Let $K^* = \min(u-1, K)$. Then the families $\mathcal{F}$ and $\mathcal{G}$ obtained as above meet conditions (iii) and (iv) of Theorem 2 if

$$K(w_1 - \tfrac{1}{2}d_1) < w_1, \quad (K - K^*)w_1 + K^*(w_2 - \tfrac{1}{2}d_2) < w_2 \quad \text{and} \quad w_2 > Kw_1. \qquad (5)$$

Proof. Since every code vector of B_1 has weight w_1 and B_1 has minimum distance d_1, each member of $\mathcal{F}$ has w_1 elements and, if any two distinct members of $\mathcal{F}$ intersect in p elements, then $2(w_1 - p) \geq d_1$, i.e., $p \leq w_1 - \frac{1}{2}d_1$. Thus any union of K or fewer members of $\mathcal{F}$ has at most $K(w_1 - \frac{1}{2}d_1)$ elements in common with any member of $\mathcal{F}$ not involved in the union. So, if the first inequality in (5) holds then $\mathcal{F}$ is K-cover-free, i.e., condition (iii) of Theorem 2 is met.

Turning to condition (iv) of Theorem 2, consider any union of K or fewer members of $\mathcal{F} \cup \mathcal{G}$ and any member of $\mathcal{G}$ not involved in the union. Suppose the union involves K_1 members of $\mathcal{F}$ and K_2 members of $\mathcal{G}$. Arguing as in the last paragraph, then there are at most $K_1 w_1 + K_2(w_2 - \frac{1}{2}d_2)$ [$= \Phi(K_1, K_2)$, say] elements common to the union and the member of $\mathcal{G}$ which does not appear in the union. Now $K_1 + K_2 \leq K$ and $0 \leq K_2 \leq K^*$, because $\mathcal{G}$ has u members one of which is not involved in the union. Hence

$$\begin{aligned} \Phi(K_1, K_2) &\leq (K - K_2)w_1 + K_2(w_2 - \tfrac{1}{2}d_2) \\ &\leq \max\{(K - K^*)w_1 + K^*(w_2 - \tfrac{1}{2}d_2), Kw_1\}. \end{aligned}$$

Since each member of $\mathcal{G}$ has w_2 elements, it is now immediate that condition (iv) of Theorem 2 is met if the last two inequalities in (5) hold. □

Given K, q, s and u, the aforesaid method of construction for $\mathcal{F}$ and $\mathcal{G}$ is suc-

cessful provided there exist binary codes $B(q, s, d_1, w_1)$ and $B(q, u, d_2, w_2)$ satisfying (5). To explore this, one needs to identify the choices of (d_1, w_1, d_2, w_2) meeting (5), and then for each such choice check the existence of $B(q, s, d_1, w_1)$ and $B(q, u, d_2, w_2)$ by verifying whether or not the conditions

$$s \le A(q, d_1, w_1) \qquad \text{and} \qquad u \le A(q, d_2, w_2) \qquad (6)$$

hold, where $A(q, d, w)$ is the largest possible N in a $B(q, N, d, w)$, given q, d and w. In particular, if $u = 1$, then $K^* = 0$ and (5) does not involve d_2. In this case, it suffices to identify the choices of (d_1, w_1, w_2) meeting (5) and for each such choice check only the existence of $B(q, s, d_1, w_1)$ by verifying the first condition in (6), with the understanding that the single code vector in $B(q, 1, d_2, w_2)$ can simply be taken as any binary vector of length q and weight w_2. Tables showing exact values of or lower bounds on $A(q, d, w)$ are useful for this purpose. Such tables appear in [2] and [10], with further significant improvements reported in [9] and the webpage http://www.win.tue.nl/~aeb/codes/Andw.html.

Example 3. This example illustrates an application of Theorem 2 by constructing a K-UDF(v, n), where K =3, v = 77 and n = 1373. We employ Theorem 2 with C obtained via Proposition 1. Then $v = mq$ and $n = M + mu = s^t + mu$. Equating these with the stipulated values of v and n, we get mq = 77, and $s^t + mu$ = 1373. These equations together with the conditions of Proposition 1, namely $t > 1$ and $K(t-1) < m$, are satisfied if m = 7, $q = s$ = 11, t = 3 and u = 6.

(a) Hence we take C as the code represented by the rows of an OA(11^3, 7, 11, 3). Then by Proposition 1, conditions (i) and (ii) of Theorem 2 are met.

(b) With $q = s$ = 11 and u = 6, we next invoke Proposition 2 to find the families of sets $\mathcal{F}$ and $\mathcal{G}$ satisfying conditions (iii) and (iv) of Theorem 2. Here K^* = 3 and the possible choices of (d_1, w_1, d_2, w_2) meeting (5) are (2, 1, 6, 4), (2, 1, 8, 4), (2, 1, 8, 5), (2, 1, 10, 5) and (2, 1, 10, 6), out of which only the first one satisfies (6). Indeed, $A(11, 2, 1) = 11$ and $A(11, 6, 4) = 6$, and starting from the associated binary codes $B(11, 11, 2, 1)$ and $B(11, 6, 6, 4)$, one can find the families $\mathcal{F}$ and $\mathcal{G}$, satisfying conditions (iii) and (iv) of Theorem 2, as given by the members $F(l)$ = $\{l\}$ (l = 0, 1, ...,10) and $G(1) = \{0, 1, 2, 3\}$, $G(2) = \{0, 4, 5, 6\}$, $G(3) = \{0, 7, 8, 9\}$, $G(4) = \{1, 4, 7, 10\}$, $G(5) = \{2, 5, 8, 10\}$, $G(6) = \{3, 6, 9, 10\}$.

With C, $\mathcal{F}$ and $\mathcal{G}$ chosen as in (a) and (b) above, Theorem 2 yields a 3-UDF(77, 1373).

With reference to Theorem 1, we now proceed to construct codes that are union distinct even without meeting (1). A method, which is shown to work for K = 2 and can potentially be extended to general K, is presented. Let s be a prime or prime power and let $\alpha_0, \alpha_1, ..., \alpha_{s-1}$ be the elements of the finite field GF(s),

$\alpha_1 = 1$ being the multiplicative identity element. Suppose $3 \le m \le s$ and define the following row vectors of order m over GF(s):

$$\rho(i) = (\alpha_0^i, \alpha_1^i, ..., \alpha_{m-1}^i) \qquad (i = 0,1,2,...). \tag{7}$$

For $2 \le t \le m$, let R be a $t \times m$ matrix with rows $\rho(i)$, $0 \le i \le t-1$, and R_0 be a $(t-1) \times m$ matrix with rows $\rho(i)$, $0 \le i \le t-2$. Define U as the $s^t \times m$ array with rows $\xi^T R$, $\xi \in S(t)$, and V as the $s^{t-1} \times m$ array with rows $\rho(t) + \mu^T R_0$, $\mu \in S(t-1)$, where $S(i)$ is the set of the s^i column vectors of order i over GF(s). It is well-known (see e.g., [3], p. 37, or [4], p. 38) that U is an OA(s^t, m, s, t) of index unity and hence, following Proposition 1, yields a code satisfying (1) if $K(t-1) < m$. Theorem 3 below shows that for $K = 2$ and odd s, under the same condition on t and m, the larger array

$$W = \begin{bmatrix} U \\ V \end{bmatrix}, \tag{8}$$

with $s^t + s^{t-1}$ rows and m columns, represents a 2-UD code.

Theorem 3. *If s is an odd prime or prime power, $t \ge 2$ and $2(t-1) < m$, then the $s^t + s^{t-1}$ rows of W, interpreted as code vectors, yield a* 2-UD *code of size $s^t + s^{t-1}$.*

Theorem 3 is proved in the appendix. With $s = m = 3$, $t = 2$ and $\alpha_l = l$ ($l = 0$, 1, 2), it yields the 2-UD code of Example 2. A more appealing application appears in Example 4 in the next section. While a 2-UD code arising from Theorem 3 may not satisfy (1) (cf. Example 2), it has more code vectors than the one given by U alone and hence yields a union distinct family with larger n via Theorem 1. Also note that if in Theorem 1, $\mathcal{F}$ is taken as a 2-cover-free family and C is obtained using Theorem 3, then the resulting $\mathcal{H}(0)$ is only partially cover-free in the sense that the subfamily of $\mathcal{H}(0)$, associated with the rows of U, is 2-cover-free.

3. Application to digital fingerprinting codes

3.1 Background

Digital fingerprinting is a technique for tracing consumers who use their multimedia contents for illegitimate purposes, such as redistribution (see [1]). Anti-collusion codes (ACCs) aim at deterring such unauthorized utilization by a coalition of users, and have been of considerable recent interest. Trappe et al. [11] introduced an attractive class of ACCs, called AND-ACCs. In order to formally define AND-ACCs, we note that the element-wise AND of a set of binary vectors $(x_{j1}, ..., x_{jv})$, where j belongs to some index set J, equals $(\prod_{j \in J} x_{j1}, ..., \prod_{j \in J} x_{jv})$.

Definition 2. *Let $X = \{(x_{j1}, ..., x_{jv}) : 1 \le j \le n\}$ be a code consisting of n binary*

code vectors each of length v. Then X is called a K-resilient AND-ACC if the element-wise ANDs of all distinct subsets of K or fewer code vectors in X are distinct, i.e., if no two distinct subsets J_1 and J_2 of $\{1,\ldots, n\}$, with both J_1 and J_2 having K or fewer elements, satisfy $\prod_{j\in J_1} x_{ji} = \prod_{j\in J_2} x_{ji}$ for each i, $1 \le i \le v$.

Let the n code vectors in X be used to watermark the digital contents of n consumers. There is a possibility that some of these consumers might collude and use their own contents to produce an illegitimate content for unauthorized redistribution. Then the watermark of this illegitimate content can be detected in the form of the element-wise AND of the code vectors used for these colluders. Therefore, if the number of colluders is K or fewer, then they can be uniquely identified whenever the element-wise ANDs of all distinct subsets of K or fewer code vectors in X are distinct, i.e., whenever X is a K-resilient AND-ACC.

A code X as in Definition 2 is denoted as a (v, n, K) AND-ACC. It involves v basis vectors, accommodates n users, and has resilience K in the sense described above. Construction of AND-ACCs is an interesting combinatorial problem which has been addressed by several researchers. Trappe et al. [11] gave a method that makes use of balanced incomplete block (BIB) designs. Kang et al. [5] proposed another approach using group-divisible designs. Yagi et al. [13] used finite geometries for obtaining AND-ACCs, while Li et al. [7] suggested a construction procedure based on cover-free families of sets. We refer to these papers for further related references. As noted by all these authors, for a given resilience K, one prefers a (v, n, K) AND-ACC with relatively large n and small v because this accommodates more users and avoids distribution of energy over a large number of basis vectors.

3.2 AND-ACCs from union distinct families

By Definition 2, a (v, n, K) AND-ACC and a K-UDF(v, n) are co-existent. To see this, it suffices to interpret the bit complements of the columns of the incidence matrix of the latter as the code vectors of the former, and vice versa. Thus the constructions in Section 2 readily yield AND-ACCs. For instance, the 3-UDF(v, n) with $v = 77$ and $n = 1373$, as obtained in Example 3, yields a (77, 1373, 3) AND-ACC. Indeed, as this example and the ones to be presented now show, given K, the AND-ACCs obtained via the constructions in Section 2 are often better than the existing ones in terms of ensuring larger n with the same or smaller v; see Remark 2 below. The same happens also in many other examples which are not reported here in order to save space.

Example 4. Suppose it desired to construct a (v, n, K) AND-ACC with $v = 60$, $n = 6972$ and $K = 2$. As $K = 2$, we consider employing Theorem 1 with C obtained via Theorem 3. The resulting 2-union distinct family will have $v = mq$ and $n = M = s^t + s^{t-1}$. Equating these with the stipulated values of v and n, we get $mq = 60$ and $s^t + s^{t-1} = 6972$. These equations together with the conditions of Theorem 3, namely s an odd prime or prime power, $t \ge 2$ and $2(t-1) < m$, are satisfied if m

$= 3, q = 20, s = 83$ and $t = 2$.

(a) Hence we use Theorem 3 with $m = 3$, $s = 83$ and $t = 2$, to obtain a 2-UD code C consisting of $M = 6972$ code vectors each of length $m = 3$.

(b) For employing Theorem 1, it remains to find a family of sets $\mathcal{F}$ which is K-UDF(q, s), where $q = 20$ and $s = 83$. Now, from Table I-B in [2], $A(20, 6, 5) \geq 84$. Thus there exists a constant weight binary code $B(20, 83, 6, 5)$ which meets the first inequality in (5) for $K = 2$ and therefore, using the same arguments as in Proposition 2, leads to a family of sets $\mathcal{F}$ which is 2-CFF(20, 83) and hence 2-UDF(20, 83).

With C and $\mathcal{F}$ chosen as in (a) and (b) above, Theorem 1 yields a 2-UDF(60, 6972) and hence a (60, 6972, 2) AND-ACC.

Example 5. This example illustrates the construction of a (v, n, K) AND-ACC with $v = 90$, $n = 6591$ and $K = 3$, employing Theorem 2 with C obtained via Proposition 1. Then one gets $v = mq$ and $n = M + mu = s^t + mu$. Equating these with the given values of v and n, we get $mq = 90$ and $s^t + mu = 6591$. These equations together with the conditions of Proposition 1, namely $t > 1$ and $K(t-1) < m$, are satisfied if $m = 10$, $q = s = 9$, $t = 4$ and $u = 3$.

(a) Hence we take C as the code represented by the rows of an OA(9^4, 10, 9, 4). Then by Proposition 1, conditions (i) and (ii) of Theorem 2 are met.

(b) With $q = s = 9$ and $u = 3$, it now remains to find the families of sets $\mathcal{F}$ and $\mathcal{G}$ satisfying conditions (iii) and (iv) of Theorem 2. Proposition 2 is used for this purpose. Here $K^* = 2$ and the possible choices of (d_1, w_1, d_2, w_2) meeting (5) are (2, 1, 6, 4), (2, 1, 8, 4) and (2, 1, 8, 5), out of which only the first one satisfies (6). Indeed, $A(9,2,1) = 9$ and $A(9,6,4) = 3$, and starting from the associated binary codes $B(9,9,2,1)$ and $B(9,3,6,4)$, one can find the families $\mathcal{F}$ and $\mathcal{G}$, satisfying conditions (iii) and (iv) of Theorem 2, as given by the members $F(l) = \{l\}$ ($l = 0, 1, \ldots, 8$) and $G(1) = \{0, 1, 2, 3\}$, $G(2) = \{0, 4, 5, 6\}$, $G(3) = \{1, 4, 7, 8\}$.

With C, $\mathcal{F}$ and $\mathcal{G}$ chosen as in (a) and (b) above, Theorem 2 yields a 3-UDF(90, 6591) and hence a (90, 6591, 3) AND-ACC.

Example 6. We now employ Theorem 2, with C obtained via Proposition 1, to construct a (v, n, K) AND-ACC where $v = 147$, $n = 29798$ and $K = 3$. As in Example 4, then we get the equations $mq = 147$ and $s^t + mu = 29798$. These equations together with the conditions of Proposition 1, namely $t > 1$ and $K(t-1) < m$, are satisfied if $m = 7$, $q = 21$, $s = 31$, $t = 3$ and $u = 1$.

(a) Hence we take C as the code represented by the rows of an OA(31^3, 7, 31, 3). Then by Proposition 1, conditions (i) and (ii) of Theorem 2 are met.

(b) With $q = 21$, $s = 31$ and $u = 1$, we next use Proposition 2 to find the families $\mathcal{F}$ and $\mathcal{G}$ satisfying conditions (iii) and (iv) of Theorem 2. Here $K^* = 0$ and (5) does not involve d_2. Among the possible choices of (d_1, w_1, w_2) meeting

(5), only the ones of the form $(6, 4, w_2)$, where $w_2 \geq 13$, satisfy the first inequality in (6); note that $A(21, 6, 4) = 31$ from Table IB in [2]. Hence if we take $\mathcal{F}$ as the family of sets given by a constant weight binary code $B(21, 31, 6, 4)$ and $\mathcal{G}$ as consisting of the only set $\{0,1,\ldots,20\}$, then conditions (iii) and (iv) of Theorem 2 are met.

With C, $\mathcal{F}$ and $\mathcal{G}$ chosen as in (a) and (b) above, Theorem 2 yields a 3-UDF(147, 29798) and hence a (147, 29798, 3) AND-ACC.

Example 7. We now use Theorem 2, with C obtained via Proposition 1, to construct a (v, n, K) AND-ACC with $v = 81$, $n = 747$ and $K = 4$. Then $mq = 81$ and $s^t + mu = 747$. These equations together with the conditions of Proposition 1, namely $t > 1$ and $K(t-1) < m$, are satisfied if $m = q = s = 9$, $t = 3$ and $u = 2$.

(a) Hence we take C as the code represented by the rows of an OA(9^3, 9, 9, 3). Then by Proposition 1, conditions (i) and (ii) of Theorem 2 are met.

(b) With $q = s = 9$ and $u = 2$, Proposition 2 is now used to find the families $\mathcal{F}$ and $\mathcal{G}$ satisfying conditions (iii) and (iv) of Theorem 2. Here $K^* = 1$ and the only choice of (d_1, w_1, d_2, w_2) meeting (5), namely (2, 1, 8, 5), satisfies (6) as well. Indeed, $A(9,2,1) = 9$ and $A(9,8,5) = 2$, and starting from the associated binary codes $B(9,9,2,1)$ and $B(9,2,8,5)$, one can find the families $\mathcal{F}$ and $\mathcal{G}$, satisfying conditions (iii) and (iv) of Theorem 2, as given by the members $F(l) = \{l\}$ ($l = 0$, 1, ...,8) and $G(1) = \{0, 1, 2, 3, 4\}$, $G(2) = \{0, 5, 6, 7, 8\}$.

With C, $\mathcal{F}$ and $\mathcal{G}$ chosen as in (a) and (b) above, Theorem 2 yields a 4-UDF(81, 747) and hence a (81, 747, 4) AND-ACC.

Remark 2. Our approach can yield AND-ACCs which are better than the existing ones in the sense of ensuring larger n with the same or smaller v, given K. For instance, in the construction in [11], n must satisfy $n \leq v(v-1)/\{(K+1)K\}$, while the values of n in our Examples 3-7 far exceed this upper bound. Similarly, in the construction in [5], $n = \{v/(K+1)\}^2$, while the AND-ACCs in Examples 3-7 have much larger values of n. Moreover, in the respective setup of these examples, the construction in [7] based on cover-free families as considered by there, yields AND-ACCs with $(v, n, K) = (77, 1331, 3)$, $(249, 6889, 2)$, $(90, 6561, 3)$, $(217, 29791, 3)$ and $(81, 729, 4)$. Our methods lead to larger n and smaller v in Examples 4 and 6, and same v but larger n in Examples 3, 5 and 7.

Appendix: Proofs of Theorems 2 and 3

Proof of Theorem 2. From the definitions of $\mathcal{H}(0)$, $\mathcal{H}(1)$,..., $\mathcal{H}(m)$, it is clear that $\mathcal{H}$ has $n = M + mu$ sets, all subsets of Ω. Consider any set H from $\mathcal{H}$ and any collection of k sets $H_1,\ldots,H_k$ from $\mathcal{H}$ such that $k \leq K$ and $H \notin \{H_1,\ldots,H_k\}$. It will suffice to show that the union of $H_1,\ldots,H_k$ does not cover H. Since $\mathcal{H} = \mathcal{H}(0) \cup \mathcal{H}(1) \ldots \cup \mathcal{H}(m)$, we can write

$$\{H_1,...,H_k\} = \Gamma_0 \cup \Gamma_1 \cup ... \cup \Gamma_m, \qquad \text{(A.1)}$$

where $\Gamma_0, \Gamma_1, ..., \Gamma_m$ are intersections of $\{H_1,...,H_k\}$ with $\mathcal{H}(0)$, $\mathcal{H}(1)$,..., $\mathcal{H}(m)$ respectively. Then $\Gamma_0, \Gamma_1, ..., \Gamma_m$ are disjoint, and if $k_0, k_1, ..., k_m$ are the numbers of sets in $\Gamma_0, \Gamma_1, ..., \Gamma_m$ respectively, then by (A.1), $k_0, k_1, ..., k_m$ satisfy

$$k_0 + k_1 + ... + k_m = k \le K. \qquad \text{(A.2)}$$

Also, as Γ_0 is a subfamily of $\mathcal{H}(0)$ consisting of k_0 members of the latter, by (2),

$$\Gamma_0 = \{ E_{j1} \cup ... \cup E_{jm} : j \in J_0 \}, \qquad \text{(A.3)}$$

where the set J_0 consists of some k_0 elements of $\{1,..., M\}$.

First, let $H \in \mathcal{H}(0)$. Then by (2), $H = E_{j^*1} \cup ... \cup E_{j^*m}$, for some j^*, where $1 \le j^* \le M$ and, in view of (A.3), $j^* \notin J_0$ because $H \notin \{H_1,..., H_k\}$. By (3) and (A.3), the k_0 sets of Γ_0 together cover E_{j^*i} for any fixed i, if and only if the union of the k_0 sets $F(\tilde{c}_{ji})$, $j \in J_0$, covers $F(\tilde{c}_{j^*i})$. Since $k_0 \le K$ and $\mathcal{F}$ is K-cover-free by condition (iii), this happens if and only if $\tilde{c}_{j^*i} \in \{\tilde{c}_{ji} : j \in J_0\}$, or equivalently, if and only if

$$c_{j^*i} \in \{c_{ji} : j \in J_0\}. \qquad \text{(A.4)}$$

But as $j^* \notin J_0$ and C has minimum distance d, (A.4) can hold for at most $k_0(m-d)$ choices of i. Thus the k_0 sets of Γ_0 ($\subseteq \mathcal{H}(0)$) together cover at most $k_0(m-d)$ of the sets E_{j^*1},..., E_{j^*m}. Also, by (3) and (4), for each i ($1 \le i \le m$), the union of the k_i sets of Γ_i ($\subseteq \mathcal{H}(i)$) has a nonempty overlap with at most one, namely E_{j^*i}, of the sets E_{j^*1},..., E_{j^*m}. Thus, by (A.1), writing $\delta_i = 1$ if $k_i > 0$ and $\delta_i = 0$ if $k_i = 0$, the union of $H_1,..., H_k$ covers at most $k_0(m-d) + \delta_1 + ... + \delta_m$ (= ψ, say) of the disjoint sets E_{j^*1},..., E_{j^*m}. Now, by (A.2),

$$\psi \le k_0(m-d) + k_1 + ... + k_m \le k_0(m-d-1) + K. \qquad \text{(A.5)}$$

If $k_0 > 0$, then by (A.2) and conditions (i) and (ii),

$$k_0(m-d-1) + K < k_0(mK^{-1} - 1) + K \le K(mK^{-1} - 1) + K = m,$$

while if $k_0 = 0$, then by condition (i), $k_0(m-d-1) + K = K < m$, so that by (A.5), $\psi < m$. As a result, the union of $H_1,..., H_k$ fails to cover at least one of the sets E_{j^*1},..., E_{j^*m}, i.e., this union does not cover H.

We next turn to the case $H \in \mathcal{H}(i)$ for some i ($1 \le i \le m$), say $H \in \mathcal{H}(1)$. Then by (4), H has empty overlap with every set in $\Gamma_2 \cup ... \cup \Gamma_m$. On the other hand, the union of the $k_0 + k_1$ sets of $\Gamma_0 \cup \Gamma_1$ cannot cover H, in view of (2)-(4), condition (iv) and the fact that $k_0 + k_1 \le K$ (vide (A.2)). This proves the result. □

Proof of Theorem 3. Given any two rows $(x_1,...,x_m)$ and $(y_1,...,y_m)$ of W, the number of coincidences between them is defined as the cardinality of the set $\{i : x_i = y_i, 1 \le i \le m\}$, i.e., it is m minus the Hamming distance between the two rows.

Lemma 1.*The number of coincidences between any two rows of W cannot exceed* (a) $t-1$, *if these are distinct rows of U,* (b) $t-2$, *if these are distinct rows of V, and* (c) t, *if one of these is a row of U and the other is a row of V.*

Proof: (a) If any two distinct rows of U have t or more coincidences, then there exists an $s^t \times t$ subarray of U where an ordered t-tuple of symbols occurs at least twice as a row. This is impossible because, as noted in Section 2, U is an $\mathrm{OA}(s^t, m, s, t)$ of index unity.

(b) By the definition of V, if (b) does not hold then there exist $\mu_1, \mu_2 \in S(t-1)$, $\mu_1 \ne \mu_2$, and a square submatrix of R_0, say $\overline{R}_0$, consisting of some $t-1$ columns of R_0 such that $(\mu_1 - \mu_2)^T \overline{R}_0$ is a null row vector. This is impossible since $\mu_1 - \mu_2 \ne 0$ while every square submatrix, of order $t-1$, of R_0 is nonsingular by (7) as $\alpha_0, \alpha_1, ..., \alpha_{m-1}$ are distinct.

(c) By the definitions of U and V, if (c) does not hold then there exist $\xi \in S(t)$ and $\mu \in S(t-1)$ such that there are $t+1$ or more coincidences between $\xi^T R$ and $\rho(t) + \mu^T R_0$, or equivalently, between $\xi^T R - \mu^T R_0$ and $\rho(t)$. Therefore, a subvector of $\rho(t)$, of order $t+1$, is in the row space of the corresponding $t \times (t+1)$ submatrix of R, because the rows of R_0 are also rows of R. This is impossible because by (7), every square submatrix, of order $t+1$, of

$$\widetilde{R} = \begin{bmatrix} R \\ \rho(t) \end{bmatrix}$$

is nonsingular as $\alpha_0, \alpha_1, ..., \alpha_{m-1}$ are distinct. □

Proof of Theorem 3 (continued). Since $2(t-1) < m$, i.e., $m \ge 2t-1$, it will suffice to prove the result for $m = 2t-1$. Then $m \ge t+1$ as $t \ge 2$, and by Lemma 1, all rows of W are distinct. Hence it is clear that no two distinct collections of rows of W, say $\{(x_{j1},...,x_{jm}) : j \in J_1\}$ and $\{(x_{j1},...,x_{jm}) : j \in J_2\}$, each with two or fewer rows, can have $\{x_{ji} : j \in J_1\} = \{x_{ji} : j \in J_2\}$ for every i $(1 \le i \le m)$, when any of the two collections has only one row or they have one row in common.

It remains to consider two collections of the form (a, b) and (x, y), where $a = (a_1,...,a_m)$, $b = (b_1,...,b_m)$, $x = (x_1,...,x_m)$ and $y = (y_1,...,y_m)$ are four distinct rows of W. If the sets $\{a_i, b_i\}$ and $\{x_i, y_i\}$ are identical for every i $(1 \le i \le m)$, then, for each i, either (i) $a_i = x_i$, $b_i = y_i$, or (ii) $a_i \ne x_i$,

$a_i = y_i$, $b_i = x_i$. Without loss of generality, suppose (i) holds for $1 \le i \le m_1$, and (ii) holds for $m_1 + 1 \le i \le m_1 + m_2$, where $m_1 + m_2 = m = 2t - 1$. Since $m_1 \le t$ and $m_2 \le t$ by (i), (ii) and Lemma 1, the pair (m_1, m_2) equals either $(t, t-1)$ or $(t-1, t)$.

First let $(m_1, m_2) = (t, t-1)$. Partition R and R_0 as $R = [R_1 \; R_2]$ and $R_0 = [R_{01} \; R_{02}]$, where R_1 and R_{01} consist of their first t columns, while R_2 and R_{02} consist of their last $t-1$ columns. Similarly, partition the row vectors $\rho(t)$, a, b, x and y as $\rho(t) = (\rho(t,1) \; \rho(t,2))$, $a = (a(1) \; a(2))$, $b = (b(1) \; b(2))$, $x = (x(1) \; x(2))$ and $y = (y(1) \; y(2))$, where $\rho(t,1)$, $a(1)$ etc. consist of the first t elements of these vectors, while $\rho(t,2)$, $a(2)$ etc. consist of their last $t-1$ elements. Then by (i) and (ii),

$$a(1) = x(1), \quad b(1) = y(1), \quad a(2) = y(2), \quad b(2) = x(2). \qquad \text{(A.6)}$$

Since $m_1 = t$, by (i) and Lemma 1, one of a and x, say a, is a row of U and the other, say x, is a row of V. Similarly, one of b and y is a row of U and the other is a row of V. But if b is a row of V and y is a row of U, then both b and x are rows of V and, by the last equation in (A.6), they have $t-1$ coincidences, which contradicts Lemma 1(b). Thus we need to consider only the situation where a and b are rows of U, and x and y are rows of V. Then by the definitions of U and V, there exist $\xi_1, \xi_2 \in S(t)$ and $\mu_1, \mu_2 \in S(t-1)$ such that

$$a = \xi_1^T R, \; b = \xi_2^T R, \qquad x = \rho(t) + \mu_1^T R_0, \qquad y = \rho(t) + \mu_2^T R_0,$$

where

$$\xi_1 \ne \xi_2, \qquad \mu_1 \ne \mu_2, \qquad \text{(A.7)}$$

as a, b, x, y are distinct. Recalling the partitioned forms of a, b, R etc., the equations in (A.6) can now be expressed as

$$\xi_1^T R_1 = \rho(t,1) + \mu_1^T R_{01}, \qquad \xi_2^T R_1 = \rho(t,1) + \mu_2^T R_{01},$$
$$\xi_1^T R_2 = \rho(t,2) + \mu_2^T R_{02}, \qquad \xi_2^T R_2 = \rho(t,2) + \mu_1^T R_{02}. \qquad \text{(A.8)}$$

As the rows of R_0 are also rows of R, we have $R_0 = QR$, for some $(t-1) \times t$ matrix Q. Thus

$$R_{01} = QR_1, \qquad R_{02} = QR_2. \qquad \text{(A.9)}$$

By (A.9), the first two equations in (A.8) yield $(\xi_1 - \xi_2)^T R_1 = (\mu_1 - \mu_2)^T Q R_1$. But by (7), R_1, being a $t \times t$ submatrix of R, is nonsingular. Therefore, $(\xi_1 - \xi_2)^T = (\mu_1 - \mu_2)^T Q$, and hence using (A.9),

$$(\xi_1 - \xi_2)^T R_2 = (\mu_1 - \mu_2)^T R_{02}. \qquad \text{(A.10)}$$

On the other hand, the last two equations in (A.8) yield

$$(\xi_1 - \xi_2)^T R_2 = (\mu_2 - \mu_1)^T R_{02}. \qquad \text{(A.11)}$$

Since s is odd, by (A.10) and (A.11), $(\mu_1 - \mu_2)^T R_{02}$ equals the null vector. But R_{02} is a $(t-1)\times(t-1)$ submatrix of R_0 and is nonsingular, by (7). Hence $\mu_1 = \mu_2$, which contradicts (A.7).

In a similar manner, a contradiction is reached when $(m_1, m_2) = (t-1, t)$. □

Acknowledgement: We thank a referee for very constructive and insightful suggestions. The work of RM was supported by the J.C. Bose National Fellowship of the Government of India and a grant from the Indian Institute of Management Calcutta.

References

[1] D. Boneh and J. Shaw (1998). Collusion-secure fingerprinting for digital data. *IEEE Trans. Inform. Theory* **44**, 1897-1905.

[2] A.E. Brouwer, J.B. Shearer, N.J.A. Sloane and W.D. Smith (1990). A new table of constant weight codes. *IEEE Trans. Inform. Theory* **36**, 1334-1380.

[3] A. Dey and R. Mukerjee (1999). *Fractional Factorial Plans*. New York: Wiley.

[4] A.S. Hedayat, N.J.A. Sloane and J. Stufken (1999). *Orthogonal Arrays: Theory and Applications*. New York: Springer-Verlag.

[5] I. Kang, K. Sinha and H.K. Lee (2006). New digital fingerprint code construction scheme using group-divisible design. *IEICE Trans. Fundamentals* **E89-A**, 3732-3735.

[6] W.H. Kautz and R.C. Singleton (1964). Nonrandom binary superimposed codes. *IEEE Trans. Inform. Theory* **10**, 363–377.

[7] Q. Li, X. Wang, Y. Li, Y. Pan and P. Fan (2009). Construction of anti-collusion codes based on cover-free families. *Sixth International Conference on Information Technology: New Generations, ITNG 2009*, Las Vegas, USA.

[8] F.J. MacWilliams and N.J.A. Sloane (1977). *The Theory of Error Correcting Codes*. Amsterdam: North Holland.

[9] R. Montemanni and D.H. Smith (2009). Heuristic algorithms for constructing binary constant weight codes. *IEEE Trans. Inform. Theory* **55**, 4651-4656.

[10] D.H. Smith, L.A. Hughes and S. Perkins (2006). A new table of constant weight codes of length greater than 28. *Electronic J. Combinatorics* **13**, # A2, 1-18.

[11] W. Trappe, M. Wu, Z.J. Wang and K.J.R. Liu (2003). Anti-collusion fingerprinting for multimedia. *IEEE Trans. Signal Processing* **51**, 1069-1087.

[12] J.C. Wang and C.F.J. Wu (1991). An approach to the construction of asymmetrical orthogonal arrays. *J. Amer. Statist. Assoc.* **86**, 450-456.

[13] H. Yagi, T. Matsushima and S. Hirasawa (2007). Improved collusion-secure codes for digital fingerprinting based on finite geometries. *IEEE Int. Conf. on System, Man and Cybernetics*, pp. 948-953.