Combinatorial properties of frameproof and traceability codes

J. N. Staddon
Bell Laboratories Research Silicon Valley
3180 Porter Drive
Palo Alto CA, 94304

D. R. Stinson and R. Wei
Department of Combinatorics and Optimization
University of Waterloo
Waterloo ON, N2L 3G1
Canada

February 24, 2000

Abstract

In order to protect copyrighted material, codes may be embedded in the content or codes may be associated with the keys used to recover the content. Codes can offer protection by providing some form of *traceability* for pirated data. Several researchers have studied different notions of traceability and related concepts in recent years. "Strong" versions of traceability allow at least one member of a coalition that constructs a "pirate decoder" to be traced. Weaker versions of this concept ensure that no coalition can "frame" a disjoint user or group of users. All these concepts can be formulated as codes having certain combinatorial properties.

In this paper, we study the relationships between the various notions, and we discuss equivalent formulations using structures such as perfect hash families. We use methods from combinatorics and coding theory to provide bounds (necessary conditions) and constructions (sufficient conditions) for the objects of interest.

1 Introduction

In this paper, we are interested in combinatorial methods to allow tracing of illegally "pirated" data. We present two scenarios to motivate the problems we consider. The first example concerns decoder boxes for decrypting broadcast messages; the second concerns variants of pay-per-view movies.

In the broadcast encryption scheme suggested by Chor, Fiat and Naor in [6], a decoder box consists of N keys, where each key takes on one of q possible values. The set of possible values for the ith key is, in general, different from the set of possible values for the jth key, if $i \neq j$. A decoder box x can be represented as an N-tuple $(x_1, \cdots, x_N)$, where $1 \leq x_i \leq q$ for $1 \leq i \leq N$. A coalition might create a pirate decoder in which, for each $1 \leq i \leq N$, the ith key is selected from one of the decoder boxes held by the coalition. From the point of view of the company that distributes decoder boxes, it would be useful to be able to identify one or more of the members of a coalition that produced a pirate decoder, once a pirate decoder is confiscated.

The second example, described in Fiat and Tassa [10], concerns pay-per-view movies. Suppose a pay-per-view movie is divided into N segments, and each segment has q possible variations. The possible variations of a segment could have the same "content", but be "marked" in some not easily detected manner. A different variation of the movie is broadcast to each subscriber. A copy of the movie, denoted x, can therefore be represented as an N-tuple $(x_1, \cdots, x_N)$, where $1 \le x_i \le q$ for $1 \le i \le N$. A coalition might try to create a pirate copy of the movie by copying segments from the versions broadcast to them, in much the same way as a coalition produced a pirate decoder box in the example described above. The cable company would like to design a scheme that enables the identification of one or more of the members of a coalition that produced a pirated movie.

1.1 Related Work

This paper studies codes with the independent parent property (IPP), traceability (TA) codes, frameproof (FP) codes and secure-frameproof (SFP) codes. IPP codes are introduced in [12]. In [12], IPP codes are studied for coalitions of pirate users of size two or less. One of the main goals of our work is to study and provide context for IPP codes when coalitions are of arbitrary size.

TA codes are discussed in [6, 17]. In [17], TA codes are studied in a more general setting, where codewords are replaced by N-subsets of a q-set. This setting is appropriate for the "decoder box" application, but not for the "pirated movie" application. More recently, traceability codes have been generalized to "dynamic traitor tracing" schemes in [10].

Frameproof codes are introduced in [5]. A stronger form of frameproof codes, secure-frameproof codes, is introduced in [16] In this paper, we demonstrate that both types of codes are weaker than IPP codes and TA codes. Hence, all of our constructions of IPP and TA codes are also examples of FP and SFP codes.

1.2 Definitions

Both of the examples from the previous section can be modeled using similar mathematical notation and definitions. As well there is weaker version of this concept ensuring that no coalition can "frame" a disjoint user or group users, introduced in [5] and [16]. Now we use a uniform notation to give several definitions, as follows.

Consider a code $\mathcal{C}$ of length N on an alphabet Q with $|Q| = q$. Then $\mathcal{C} \subseteq Q^N$ and we will call it an (N, n, q)-code if $|\mathcal{C}| = n$. The elements of $\mathcal{C}$ called *codewords*; each codeword $x = (x_1, \cdots, x_N)$, where $x_i \in Q$, $1 \le i \le N$.

For any subset of codewords $\mathcal{C}_0 \subseteq \mathcal{C}$, we define the set of *descendants* of $\mathcal{C}_0$, denoted $\mathsf{desc}(\mathcal{C}_0)$ by

$$\mathsf{desc}(\mathcal{C}_0) = \{x \in Q^N : x_i \in \{a_i : a \in \mathcal{C}_0\}, 1 \le i \le N\}.$$

The set $\mathsf{desc}(\mathcal{C}_0)$ consists of the N-tuples that could be produced by a coalition holding the codewords in the set $\mathcal{C}_0$.

Now, let w be a positive integer. For a code $\mathcal{C}$, define the *w-descendant code* of $\mathcal{C}$, denoted $\mathsf{desc}_w(\mathcal{C})$, as follows:

$$\mathsf{desc}_w(\mathcal{C}) = \bigcup_{\mathcal{C}_0 \subseteq \mathcal{C}, |\mathcal{C}_0| \le w} \mathsf{desc}(\mathcal{C}_0).$$

The set $\mathsf{desc}_w(\mathcal{C})$ consists of the N-tuples that could be produced by some coalition of size at most w.

We now give the following definitions concerning traceability properties of codes.

Definition 1.1 Suppose $\mathcal{C}$ is an (N, n, q)-code and $w \geq 2$ is an integer. Let $\mathcal{C}_i \subseteq \mathcal{C}, i = 1, 2, \cdots, t$, be all the subsets of $\mathcal{C}$ such that $|\mathcal{C}_i| \leq w$. (Hence $t = \sum_{j=1}^{w} \binom{n}{j}$.)

1. $\mathcal{C}$ is a w-FP (*frameproof*) code provided that for all $x \in \mathsf{desc}_w(\mathcal{C})$, $x \in \mathsf{desc}(\mathcal{C}_i) \cap \mathcal{C}$ implies $x \in \mathcal{C}_i$.

2. $\mathcal{C}$ is a w-SFP (*secure-frameproof*) code provided that for all $x \in \mathsf{desc}_w(\mathcal{C})$, $x \in \mathsf{desc}(\mathcal{C}_i) \cap \mathsf{desc}(\mathcal{C}_j)$ implies that $\mathcal{C}_i \cap \mathcal{C}_j \neq \emptyset$, where $i \neq j$.

3. $\mathcal{C}$ is a w-IPP (*identifiable parent property*) code provided that for all $x \in \mathsf{desc}_w(\mathcal{C})$, it holds that
$$\bigcap_{\{i : x \in \mathsf{desc}(\mathcal{C}_i)\}} \mathcal{C}_i \neq \emptyset.$$

4. For $x, y \in Q^N$, define $I(x, y) = \{i : x_i = y_i\}$. $\mathcal{C}$ is a w-TA (*traceability*) code provided that, for all $x \in \mathsf{desc}_w(\mathcal{C})$, $x \in \mathsf{desc}(\mathcal{C}_i)$ implies that there is at least one codeword $y \in \mathcal{C}_i$ such that $|I(x, y)| > |I(x, z)|$ for any $z \in \mathcal{C} \backslash \mathcal{C}_i$.

The "meaning" of the above definitions is as follows:

- A code is w-frameproof if no coalition of size at most w can frame another user not in the coalition by producing the codeword held by that user. A code is w-secure frameproof if no coalition of size at most w can frame a disjoint coalition of size at most w by producing an N-tuple that could have been produced by the second coalition. w-frameproof and w-secure-frameproof codes are discussed in [5, 17, 16] for binary codes. Our definitions here extend these concepts to the non-binary case. Note that, in our model, we do not allow "unreadable" bits in the N-tuples (see [5, 16]).

- A code has the w-identifiable parent property if no coalition of size at most w can produce an N-tuple that cannot be traced back to at least one member of the coalition.

- w-traceability codes are also w-IPP codes (see Lemma 1.3 below). The advantage of the w-TA property is that it allows an efficient (i.e., linear-time) algorithm to determine an identifiable parent.

Example 1.2 We present a $(5, 16, 4)$ 2-TA code, which is an application of a general construction to be presented later (see Theorem 4.5):

$$
\begin{array}{rccccc}
\mathbf{c}_1 = & 1 & 1 & 1 & 1 & 1 \\
\mathbf{c}_2 = & 1 & 2 & 2 & 2 & 2 \\
\mathbf{c}_3 = & 1 & 3 & 3 & 3 & 3 \\
\mathbf{c}_4 = & 1 & 4 & 4 & 4 & 4 \\
\mathbf{c}_5 = & 2 & 1 & 2 & 3 & 4 \\
\mathbf{c}_6 = & 2 & 2 & 1 & 4 & 3 \\
\mathbf{c}_7 = & 2 & 3 & 4 & 1 & 2 \\
\mathbf{c}_8 = & 2 & 4 & 3 & 2 & 1 \\
\mathbf{c}_9 = & 3 & 1 & 4 & 2 & 3 \\
\mathbf{c}_{10} = & 3 & 2 & 3 & 1 & 4 \\
\mathbf{c}_{11} = & 3 & 3 & 2 & 4 & 1 \\
\mathbf{c}_{12} = & 3 & 4 & 1 & 3 & 2 \\
\mathbf{c}_{13} = & 4 & 1 & 3 & 4 & 2 \\
\mathbf{c}_{14} = & 4 & 2 & 4 & 3 & 1 \\
\mathbf{c}_{15} = & 4 & 3 & 1 & 2 & 4 \\
\mathbf{c}_{16} = & 4 & 4 & 2 & 1 & 3
\end{array}
$$

□

1.3 Fundamental Results

It is easy to see that w-IPP implies w-SFP and w-SFP implies w-FP. The following lemma shows that w-TA implies w-IPP.

Lemma 1.3 *A w-TA code is a w-IPP code.*

Proof. Suppose $\mathcal{C}$ is a w-TA code. If $x \in \mathsf{desc}_w(\mathcal{C})$, then there is a subset $\mathcal{C}_i \subseteq \mathcal{C}$, where $|\mathcal{C}_i| = w$, such that $x \in \mathsf{desc}(\mathcal{C}_i)$. Let $y \in \mathcal{C}_i$ such that $|I(x,y)| \geq |I(x,z)|$ for all $z \in \mathcal{C}_i$. Thus $|I(x,y)| \geq |I(x,z)|$ for any $z \in \mathcal{C}$ by the definition of a w-TA code. We will show that, for any $\mathcal{C}_j \subseteq \mathcal{C}$ with $|\mathcal{C}_j| \leq w$, $x \in \mathsf{desc}(\mathcal{C}_j)$ implies $y \in \mathcal{C}_j$. In fact, if $y \notin \mathcal{C}_j$, then there is $w \in \mathcal{C}_j$ such that $|I(x,w)| > |I(x,y)|$ by the definition of a w-TA code. This contradicts the fact that $|I(x,y)| \geq |I(x,z)|$ for any $z \in \mathcal{C}$. □

The following example shows that a code having the IPP property does not necessarily have the TA property.

Example 1.4 A 4-IPP $(3,4,4)$-code which is not a 2-TA code.

Let $\mathcal{C} = \{011, 123, 211, 332\}$. Then $\mathcal{C}$ is 4-IPP, since the symbols in the first position of all the codewords are different. But it is not a 2-TA code. For example, let $x = 111$. Then x is a descendant of $\{123, 011\}$. However, $|I(x \cap 123)| = 1$ and $|I(x \cap 011)| = |I(x \cap 211)| = 2$. Thus the code is not a 2-TA code. □

Remark. Several explicit constructions for 2-IPP codes are given in [12]. It can be verified that the majority of these codes are not 2-TA codes. So the above example is not unusual.

Remark. It is easy to see that there are w-FPC codes that are not w-IPP codes. For example, constructions are given in [16] for w-FPC codes on a binary alphabet, for all

$w \geq 2$. However, we will show that a w-IPP code cannot exist on an alphabet of size less than w.

The purpose of all the types of codes studied in this paper is to discourage a possible coalition from constructing illegal N-tuples. If an illegal N-tuple is constructed by a coalition of size at most w, then it is possible to identify at least one of the traitors if the code is w-IPP or w-TA. The following lemma shows that we cannot expect to identify all the traitors, except for certain "trivial" codes.

Lemma 1.5 *Suppose $\mathcal{C}$ is any (N, n, q) code with $n > q$. Then there exist three codewords y, z, z' and $x \in Q^N$ such that $x \in \mathsf{desc}(\{y, z\}) \cap \mathsf{desc}(\{y, z'\})$.*

Proof. There exists a coordinate k such that $z_k = z'_k \neq y_k$. For convenience, assume that $k = 1$. Then define $x = (x_1, \cdots, x_N) \in Q^N$ as follows:

$$\begin{aligned} x_1 &= z_1 \quad (= z'_1) \\ x_i &= y_i \quad \text{for } 2 \leq i \leq N. \end{aligned}$$

Clearly, $x \in \mathsf{desc}(\{y, z\}) \cap \mathsf{desc}(\{y, z'\})$. □

Note that there is always a trivial (N, q, q) code which is "totally traceable": the codewords are $(1, \ldots, 1)$, ... , $(q, \ldots, q)$. In this paper, we are interested only in codes with $n > q$.

We now prove another impossibility result, which shows that w-IPP codes cannot exist for certain parameter situations. This result is a generalization of [16, Theorem 2.1].

Lemma 1.6 *Suppose $\mathcal{C}$ is any (N, n, q) code, and $n - 1 \geq w \geq q$. Then $\mathcal{C}$ is not a w-IPP code.*

Proof. Let $z^1, \cdots, z^{w+1} \in \mathcal{C}$. For $1 \leq i \leq N$, let y_i be chosen such that $|\{j : z_i^j = y_i\}| \geq 2$. (This can be done using the pigeonhole principle, because $w + 1 > q$.) Then let $y = (y_1, \cdots, y_N)$. Now, it is easy to see that $y \in \mathsf{desc}(\{z^1, \cdots, z^{w+1}\} \backslash \{z^j\})$ for any j, $1 \leq j \leq w + 1$. Hence $\mathcal{C}$ is not w-IPP. □

Finally, we notice that the case of $w \geq N$ is very restrictive, as shown in the following lemma.

Lemma 1.7 *Suppose $\mathcal{C}$ is an (N, n, q) w-FP code, where $w \geq N$. Then $q \geq n$.*

Proof. If $q < n$, then we can show that $\mathcal{C}$ is not w-FP. Let $x \in \mathcal{C}$. Since $q < n$, there is a codeword $x^1 \neq x$ such that $x_1 = x_1^1$. In a similar way, we can find $x^2, \cdots, x^N$ such that $x^i \neq x$ and $x_i^i = x_i$ for $1 \leq i \leq N$. Thus $x \in \mathsf{desc}(x^1, \cdots, x^N)$. □

As mentioned above, codes with $q \geq n$ exist trivially.

2 Connections between Hash Families and Traceability Codes

Perfect hash families have undergone considerable study due to their applications in information retrieval; see [7] for an extensive survey. More recently, perfect hash families and related structures such as separating hash families (see [16]) have found applications in cryptography. We will discuss some of these applications in this section; other applications are given in [2].

Definition 2.1 Let $n \geq m$. An (n,m)*-hash function* is a function $h : A \to B$, where $|A| = n$ and $|B| = m$. An (n,m)*-hash family* is a finite set $\mathcal{H}$ of (n,m)-hash functions such that $h : A \to B$ for each $h \in \mathcal{H}$, where $|A| = n$ and $|B| = m$. We use the notation $\mathrm{HF}(N; n, m)$ to denote an (n,m)-hash family with $|\mathcal{H}| = N$.

Definition 2.2 Let n, m and w be integers such that $n \geq m \geq w \geq 2$. An (n,m,w)*-perfect hash family* is an (n,m)-hash family, $\mathcal{H}$, such that for any $X \subseteq A$ with $|X| = w$, there exists at least one $h \in \mathcal{H}$ such that $h|_X$ is injective. We use the notation $\mathrm{PHF}(N; n, m, w)$ to denote an (n,m,w)-perfect hash family with $|\mathcal{H}| = N$.

Definition 2.3 Let n, m, w_1 and w_2 be positive integers such that $n \geq m$. An (n, m, w_1, w_2)*-separating hash family* is an (n,m)-hash family, $\mathcal{H}$, such that for any $X_1, X_2 \subseteq A$ with $|X_1| = w_1$, $|X_2| = w_2$ and $X_1 \cap X_2 = \emptyset$, there exists at least one $h \in \mathcal{H}$ such that $\{h(x) : x \in X_1\} \cap \{h(x) : x \in X_2\} = \emptyset$. We use the notation $\mathrm{SHF}(N; n, m, w_1, w_2)$ to denote an $(n, m, , w_1, w_2)$-separating hash family with $|\mathcal{H}| = N$.

We can depict a (N, n, q)-code, $\mathcal{C}$, as an $n \times N$ matrix on q symbols, where each row of the matrix corresponds to one of the codewords. Similarly, we can represent an $\mathrm{HF}(N; n, m)$, $\mathcal{H}$, as an $N \times n$ matrix on m symbols, where each row of the matrix corresponds to one of the functions in $\mathcal{H}$.

Given an (N, n, q)-code $\mathcal{C}$, we define $\mathcal{H}(\mathcal{C})$ to be the $\mathrm{HF}(N; n, q)$ whose matrix representation is C^T. Thus if $\mathcal{C} = \{x^1, x^2, \cdots, x^n\}$ and $1 \leq j \leq N$, then the hash function $h_j \in \mathcal{H}(\mathcal{C})$ is defined by the rule $h_j(i) = x^i_j$, $1 \leq i \leq n$.

Connections between separating and perfect hash families on the one hand, and codes with traceability properties, on the other hand, have been pointed out in several previous papers. We summarize previous results of this nature now.

Theorem 2.4 *[16] A (N, n, q)-code, $\mathcal{C}$, is a w-FP code if and only if $\mathcal{H}(\mathcal{C})$ is an $\mathrm{SHF}(N; n, q, w, 1)$.*

Theorem 2.5 *[16] A (N, n, q)-code, $\mathcal{C}$, is a w-SFP code if and only if $\mathcal{H}(\mathcal{C})$ is an $\mathrm{SHF}(N; n, q, w, w)$, where $n \geq 2w$.*

Theorem 2.6 *[12, Lemma 1] A (N, n, q)-code, $\mathcal{C}$, is a 2-IPP code if and only if $\mathcal{H}(\mathcal{C})$ is simultaneously a $\mathrm{PHF}(N; n, q, 3)$ and an $\mathrm{SHF}(N; n, q, 2, 2)$.*

We note that Theorems 2.4 and 2.5 are proved for binary alphabets in [17]. The extension to nonbinary alphabets is straightforward.

A relationship between w-IPP codes and perfect and separating hash families is given in the following theorem.

Theorem 2.7 *Suppose $\mathcal{C}$ is an (N, n, q) w-IPP code. Then we have the following.*

1. *$\mathcal{H}(\mathcal{C})$ is a $\mathrm{PHF}(N; n, q, w+1)$ if $n \geq w + 1$.*
2. *$\mathcal{H}(\mathcal{C})$ is an $\mathrm{SHF}(N; n, q, w, w)$ if $n \geq 2w$.*

Proof. Suppose $\mathcal{H}(\mathcal{C})$ is not a $\mathrm{PHF}(N; n, q, w+1)$. Then there exists $w+1$ codewords in $\mathcal{C}$ such that in the ith position, for $1 \leq i \leq N$, there are at least two codewords having the same symbol x_i. Let $x = (x_1, x_2, \cdots, x_N)$. Then it is easy to check that any w of these $w + 1$ codewords can produce x. Thus $\mathcal{C}$ does not have the w-IPP property.

The second conclusion follows from Theorem 2.5. $\square$

Corollary 2.8 *An (N, n, q) w-IPP code does not exist if $q \leq w$.*

We cannot prove that the converse of Theorem 2.7 holds for $w > 2$. However, we can obtain w-IPP codes from certain perfect hash families, as follows.

Theorem 2.9 *An (N, n, q)-code, $\mathcal{C}$, is a w-IPP code if $\mathcal{H}(\mathcal{C})$ is a PHF$(N; n, q, \lfloor (w+2)^2/4 \rfloor)$.*

Proof. Suppose there is an $x \in \mathsf{desc}_w(\mathcal{C})$ such that

$$\bigcap_{\{i : x \in \mathsf{desc}(\mathcal{C}_i), |\mathcal{C}_i| \leq w\}} \mathcal{C}_i = \emptyset.$$

Define

$$D = \{\mathcal{C}_i : x \in \mathsf{desc}(\mathcal{C}_i), |\mathcal{C}_i| \leq w\}.$$

D is the set of all coalitions of size at most w, that could have produced x. Let

$$\alpha = \min\{|D'| : D' \subseteq D, \bigcap_{\mathcal{C}_i \in D'} \mathcal{C}_i = \emptyset\}.$$

Without loss of generality, let $D_1 = \{\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_\alpha\}$ be a collection of α distinct coalitions in D, such that no codeword is common to all the coalitions in D_1 and each coalition in D_1 can produce x. Suppose

$$\bigcup_{\mathcal{C}_i \in D_1} \mathcal{C}_i = \{y^1, y^2, \cdots, y^\beta\}.$$

Then, for $1 \leq i \leq \alpha$, there is a codeword

$$y^{k_i} \in \bigcap_{j=1, j \neq i}^{\alpha} \mathcal{C}_j$$

such that $y^{k_i} \notin \mathcal{C}_i$, since α is the minimum number of coalitions having an empty intersection. Thus there are at most $w - (\alpha - 1)$ codewords in $\mathcal{C}_i$ that are not included in the set $\{y^{k_1}, y^{k_2}, \cdots, y^{k_\alpha}\}$. Hence

$$\beta \leq \alpha + \alpha(w - \alpha + 1) = (w + 2 - \alpha)\alpha.$$

α is an integer, so it follows that

$$\beta \leq \left\lfloor \frac{(w+2)^2}{4} \right\rfloor.$$

Now, since $\mathcal{H}(\mathcal{C})$ is a PHF$(N; n, q, \beta)$, there exists a $j, 1 \leq j \leq N$, such that the elements y^i_j are all distinct for $1 \leq i \leq \beta$. Now consider x. There exists an r, $1 \leq r \leq \beta$, such that $x_j = y^r_j$. The codeword y^r cannot be in every coalition in D_1 because no codeword is in every coalition in D_1. Let $\mathcal{C}_\ell$ be a coalition in D_1 that doesn't contain y^r. Since none of the other codewords in $\bigcup_{\mathcal{C}_i \in D_1} \mathcal{C}_i$ agree with y^r on the jth index, $x \notin \mathsf{desc}(\mathcal{C}_\ell)$, a contradiction. $\square$

In order for Theorem 2.9 to be applied, it must be the case that $q \geq \lfloor (w+2)^2/4 \rfloor$. We proved earlier that a w-IPP code does not exist if $q \leq w$. This leaves open the question of the existence of w-IPP codes for $w < q < \lfloor (w+2)^2/4 \rfloor$.

The relationships between the different structures we have defined are summarized in Figure 1. (Note that the term "CFF" (cover-free family) will be defined in the next section, and "CFC" (cover-free code) will be defined in Section 4.)

Figure 1: Relationships among different types of codes and hash families

$$
\begin{array}{ccccc}
 & & \text{code with } d > N(1-1/w^2) & & \\
 & & \Downarrow & & \\
 & & (w, 1-1/w)\text{-CFC} & & \\
 & & \Downarrow & & \\
 & & w\text{-TA} & & \\
 & & \Downarrow & & \\
\left\lfloor \frac{(w+2)^2}{4} \right\rfloor\text{-PHF} & \Longrightarrow & w\text{-IPP} & \Longrightarrow & (w+1)\text{-PHF} \\
 & & \Downarrow & & \\
(w,w)\text{-SHF} & \Longleftrightarrow & w\text{-SFP} & & \Downarrow \\
 & & \Downarrow & & \\
 & & w\text{-FP} & \Longleftrightarrow & (w,1)\text{-SHF} \\
 & & \Downarrow & & \\
 & & w\text{-CFF} & &
\end{array}
$$

Key

CFC	cover-free code
CFF	cover-free family
PHF	perfect hash family
SHF	separating hash family
TA	traceability code
IPP	identifiable parent property code
FP	frameproof code
SFP	secure frameproof code

3 Necessary Conditions

In our codes, we want the value of n to be as large as possible, given values for q, N and w. In this section we discuss some upper bounds on the value of n, which yield necessary conditions for the existence of the codes.

3.1 Bounds from Cover-Free Families

In order to obtain our first bound, we employ a type of set system called a cover-free family.

Definition 3.1 A *w-cover-free family* is a pair $(X, \mathcal{B})$, where X is a set of size q and $\mathcal{B}$ is a set of subsets of X, such that for any $\mathcal{B}_0 \subseteq \mathcal{B}$ with $|\mathcal{B}_0| \leq w$ and for any $A \in \mathcal{B} \backslash \mathcal{B}_0$, it holds that

$$A \not\subseteq \bigcup_{B \in \mathcal{B}_0} B.$$

A w-cover-free family will be denoted as w-CFF(q, n) if $|\mathcal{B}| = n$. A w-cover-free family is said to be *N-uniform* if $|B| = N$ for all $B \in \mathcal{B}$.

Cover-free families have been studied extensively in combinatorics, and also have various applications in computer science, such as group testing algorithms to name one example. There are numerous cryptographic applications of cover-free families; see, for example, [5, 6, 8, 11, 13, 15, 17, 16]. Here we consider the relationship between cover-free families and w-FP codes.

Lemma 3.2 *Suppose $\mathcal{C}$ is an (N, n, q) code on an alphabet Q. Define $X = \{1, \cdots, N\} \times Q$, and for each codeword $c = (c_1, \cdots, c_N) \in \mathcal{C}$, define an N-subset of X as follows:*

$$B_c = \{(i, c_i) : 1 \leq i \leq N\}.$$

Finally, define $\mathcal{B} = \{B_c : c \in \mathcal{C}\}$. Then the set system $(X, \mathcal{B})$ is an N-uniform w-CFF(qN, n) if and only if $\mathcal{C}$ is a w-FP code.

Proof. Suppose $(X, \mathcal{B})$ is an N-uniform w-CFF(qN, n). Let $\mathcal{C}_i = \{c_{i_1}, c_{i_2}, \cdots, c_{i_w}\}$ be any w-subset of $\mathcal{C}$. For any N-tuple $x \in \mathsf{desc}(\mathcal{C}_i) \cap \mathcal{C}$, it must be the case that $B_x \subseteq \cup_{j=1}^{w} B_{i_j}$. By the definition of w-CFF, $B_x = B_{i_j}$ for some $j, 1 \leq j \leq w$. This means that $x \in \mathcal{C}_i$. Thus $\mathcal{C}$ is a w-FP code.

Conversely, suppose $\mathcal{C}$ is a w-FP code. For any $\mathcal{B}' = \{B_{c_1}, B_{c_2}, \cdots, B_{c_w}\} \subseteq \mathcal{B}$ and $B_x \in \mathcal{B} \backslash \mathcal{B}'$, we must have $B_x \not\subseteq \cup_{B \in \mathcal{B}'} B$. Otherwise, we will have $x \in \mathsf{desc}(\mathcal{C}_i) \cap \mathcal{C}$, where $\mathcal{C}_i = \{c_1, c_2, \cdots, c_w\}$, but $x \notin \mathcal{C}_i$, which contradicts the definition of w-FP code. □

There is an upper bound for uniform cover-free families, proved by Erdös, Frankl and Füredi in [9], as follows.

Theorem 3.3 *In any N-uniform w-CFF(q, n), it holds that*

$$n \leq \frac{\binom{q}{t}}{\binom{N-1}{t-1}},$$

where $t = \lceil N/w \rceil$.

Theorem 3.3, together with Lemma 3.2, gives an upper bound on n for all the structures in Figure 1. For example, we can obtain the following bound, which is similar to [17, Theorem 5.5].

Theorem 3.4 *In an (N, n, q) w-TA, w-IPP, w-SFP, or w-FP code, the following bound holds:*

$$n \leq \frac{\binom{qN}{t}}{\binom{N-1}{t-1}},$$

where $t = \lceil \frac{N}{w} \rceil$.

We can simplify the bound as follows.

$$\begin{aligned}
\frac{\binom{qN}{\lceil \frac{N}{w} \rceil}}{\binom{N-1}{\lceil \frac{N}{w} \rceil - 1}} &= \frac{w\binom{qN}{\lceil \frac{N}{w} \rceil}}{\binom{N}{\lceil \frac{N}{w} \rceil}} \\
&< w\left(\frac{qN - \frac{N}{w}}{N - \frac{N}{w}}\right)^{\frac{N}{w}} \\
&= w\left(\frac{qw-1}{w-1}\right)^{\frac{N}{w}} \\
&\approx wq^{\frac{N}{w}}.
\end{aligned}$$

Roughly speaking, n is bounded above by $wq^{\frac{N}{w}}$.

3.2 Bounds from Separating Hash Families

Since we have proved some relationships between codes and separating hash families, bounds on SHF will also give us bounds on codes. In this subsection, we investigate some bounds on SHF. First, we state and prove a simple lemma from [4].

Lemma 3.5 *Suppose A is an $m \times n$ matrix on q symbols. If $n > q^m$, then there are at most $q^m - 1$ non-repeated columns in A .*

Proof. There are at most q^m different columns in A, and at least one column is repeated since $n > q^m$. □

For a matrix A, we use R_A to denote the set of all non-repeated columns. The following theorem gives a bound for SHF$(N; n, q, w, 1)$; the proof is identical to the proof of a bound for PHF due to Blackburn and Wild [4].

Theorem 3.6 *Suppose there is an SHF$(N; n, q, w, 1)$. If $w \geq 2$ and $n > w(q^e - 1)$, then $N > we$.*

Proof. Suppose $N \leq we$. Let $A = (a_{i,j})$ be the $N \times n$ matrix obtained from the SHF$(N; n, q, w, 1)$. Divide A into w submatrices of size $e \times n$ and denote them as $A_1, A_2, \cdots, A_w$. Thus $A_1 = (a_{i,j})$, $1 \leq i \leq e$, $1 \leq j \leq n$; $A_2 = (a_{i,j})$, $e+1 \leq i \leq 2e$, $1 \leq j \leq n$; etc.

From Lemma 3.5 we have

$$\left| \bigcup_{i=1}^{w} R_{A_i} \right| \leq w(q^e - 1).$$

Since $n > w(q^e - 1)$, there is at least one column, say column j_0, which is disjoint from $\cup_{i=1}^{w} R_{A_i}$. This is a column N-tuple, say $(a_1, a_2, \cdots, a_N)^T$. For each i such that $1 \le i \le w$, we can find a column in A_i, say column j_i, which is the same as $(a_{(i-1)e+1}, a_{(i-1)e+2}, \cdots, a_{ie})^T$. Let $W = \{j_1, \cdots, j_w\}$. Then there is no hash function separating $\{j_0\}$ and W. ∎

Letting $e = \lceil \frac{N}{w} \rceil$, the above bound on separating hash families gives us a bound which is slightly stronger than Theorem 3.4.

Theorem 3.7 *In an (N, n, q) w-TA, w-IPP, w-SFP, or w-FP code, the following bound holds:*

$$n \le w \left(q^{\lceil \frac{N}{w} \rceil} - 1 \right).$$

Since a w-PHF is a $(w-1, 1)$-SHF, we also have the following corollary that was first proved in [4].

Corollary 3.8 *[4, Theorem 1] Suppose there is a PHF$(N; n, q, w)$. If $w \ge 3$ and $n > (w-1)(q^e - 1)$, then $N > (w-1)e$.*

We now obtain a bound for SHF$(N; n, q, w, w)$, as follows.

Theorem 3.9 *Suppose there is an SHF$(N; n, q, w, w)$. If $w \ge 2$ and $n > q^e + 2w - 2$, then $N > we$.*

Proof. Suppose $N \le we$. Let A and $A_1, \cdots, A_w$ be the same as in the proof of Theorem 3.6. Then there are two identical columns in A_1, say columns i_1 and j_1. Let A_2' be the matrix formed by deleting columns i_1 and j_1 from A_2. Then there are two identical columns, say columns i_2 and j_2, in A_2'. In general, for $h \le w$, we can find identical columns, say columns i_h and j_h, in A_h, where $i_h, j_h \notin \{i_1, j_1, \cdots, i_{h-1}, j_{h-1}\}$. It is easily seen that there is no hash function separating $\{i_1, \cdots, i_w\}$ from $\{j_1, \cdots, j_w\}$, a contradiction. ∎

Theorem 3.9 gives us a bound on w-SFP (and thus w-IPP and w-TA) as follows:

Theorem 3.10 *In an (N, n, q) w-TA, w-IPP or w-SFP code, the following bound holds:*

$$n \le q^{\lceil \frac{N}{w} \rceil} + 2w - 2.$$

This bound is stronger than the previous bound obtained from cover-free families, Theorem 3.7 (of course Theorem 3.7 provides a bound for w-FP codes while Theorem 3.10 does not).

The above bound for w-FP, w-SFP, w-IPP and w-TA codes are the best bounds known for $w > 2$. For 2-IPP codes, a stronger bound was shown in [12]:

Theorem 3.11 *In an (N, n, q) 2-IPP code, it holds that $n \le 3q^{\lceil N/3 \rceil}$.*

We can use the same techniques to derive a new upper bound for (w_1, w_2)-SHF.

Theorem 3.12 *Suppose there is an SHF$(N; n, q, w_1, w_2)$, where $w_1 > w_2 \ge 2$. If*

$$n > \max\{(w_1 - w_2 + 1)(q^e - 1), q^e + w_1 + w_2 - 2\},$$

then $N > w_1 e$.

Proof. Suppose there exists an SHF$(N; n, q, w_1, w_2)$, where all the given conditions hold, but $N \le w_1 e$. Let $A = (a_{i,j})$ be the $N \times n$ matrix obtained from the SHF$(N; n, q, w_1, w_2)$. Divide A into two submatrices of sizes $(w_1 - w_2 + 1)e \times n$ and $(w_2 - 1)e \times n$, and denote them as A_1 and A_2, respectively. By Theorem 3.6, there is a set Z of $(w_1 - w_2 + 1)$ columns of A_1, and an additional column z of A_1, such that no hash function in A_1 separates Z from $\{z\}$.

Now, delete the columns $Z \cup \{z\}$ from A_2. By Theorem 3.9, there are two disjoint sets of $w_2 - 1$ columns of A_2, say Y_1 and Y_2, such that no hash function in A_2 separates Y_1 from Y_2.

Thus the hash functions in A cannot separate $Y_1 \cup Z$ and $Y_2 \cup \{z\}$, a contradiction. □

4 Sufficient Conditions

In this section, we consider sufficient conditions for the existence of the codes of interest. These take two forms:

- explicit constuctions, utilizing error-correcting codes, and
- nonconstructive existence results that utilize the probabilistic method [1].

4.1 Constructions Using Error-correcting Codes

In [13], a stronger form of CFF was introduced in connection with a broadcast encryption technique. It was also remarked in [13] that this stronger form of CFF could be used to construct traitor tracing schemes. We pursue this theme now, adapting their definition to the setting of codes that we use in this paper.

Definition 4.1 Suppose that $\mathcal{C}$ is an (N, n, q) code. For any subset $\mathcal{C}' \subseteq \mathcal{C}$ and any $x \in Q^N$, define $I(x, \mathcal{C}') = \{i : x_i = y_i \text{ for some } y \in \mathcal{C}'\}$. Then $\mathcal{C}$ is called (w, α)*-cover-free code*, denoted (w, α)-CFC, if $|I(z, \mathcal{C}')| < (1-\alpha)N$ for any w-subset $\mathcal{C}' \subseteq \mathcal{C}$ and any $z \in \mathcal{C} \backslash \mathcal{C}'$.

The following theorem shows that a CFC code is also a TA code.

Theorem 4.2 *Suppose that an (N, n, q) code $\mathcal{C}$ is a $(w, 1 - 1/w)$-CFC. Then $\mathcal{C}$ is a w-TA code.*

Proof. Suppose $x \in \mathsf{desc}(\mathcal{C}')$, where $\mathcal{C}' \subseteq \mathcal{C}$, $|\mathcal{C}'| = w$. Then there exists $y \in \mathcal{C}'$ such that $|I(x, y)| \ge w/N$. On the other hand, for any $z \in \mathcal{C} \backslash \mathcal{C}'$, $|I(x, z)| \le |I(z, \mathcal{C}')| < w/N$, since $I(z, x) \subseteq I(z, \mathcal{C}')$. Thus $\mathcal{C}$ is a w-TA code. □

Codes with large minimum distance are CFC codes. We prove the following simple result.

Theorem 4.3 *Suppose that $\mathcal{C}$ is an (N, n, q)-code having minimum distance $d > N(1 - 1/w^2)$. Then $\mathcal{C}$ is a $(w, 1 - 1/w)$-CFC.*

Proof. Suppose $\mathcal{C}' \subseteq \mathcal{C}$, $|\mathcal{C}'| = w$ and $z \in \mathcal{C} \backslash \mathcal{C}'$. Then for any $y \in \mathcal{C}'$, we have

$$|I(z, y)| < N - N(1 - 1/w^2) = N/w^2.$$

Thus

$$|I(z, \mathcal{C}')| < N/w = (1 - \alpha)N,$$

where $\alpha = 1 - 1/w$. □

The following result, stated in [6], is an immediate corollary of the two preceding theorems.

Theorem 4.4 *Suppose that $\mathcal{C}$ is an (N, n, q)-code having minimum hamming distance $d > N(1 - 1/w^2)$. Then $\mathcal{C}$ is a w-TA code.*

In [17], an explicit construction for w-TA codes was presented that used orthogonal arrays. This construction can be viewed as a corollary of Theorem 4.4 using Reed-Solomon codes. Independently, Reed-Solomon codes were used in [12] to construct 2-IPP codes in a similar fashion. These results are all contained in the following more general theorem.

Theorem 4.5 *Suppose N, q and w are given, with q a prime power and $N \leq q + 1$. Then there exists an (N, n, q) w-TA code in which $n = q^{\lceil N/w^2 \rceil}$.*

Proof. Suppose N, q and w are given, with q a prime power and $N \leq q+1$. Let $t = \lceil N/w^2 \rceil$. Then there exists a q-ary Reed-Solomon code of length N and dimension t, say $\mathcal{C}$. $\mathcal{C}$ is an (N, q^t, q)-code with minimum hamming distance $d = N - t + 1$. It is easy to check that $d > N(1 - 1/w^2)$, so Theorem 4.4 can be applied. Therefore $\mathcal{C}$ is an (N, n, q) w-TA code in which $n = q^{\lceil N/w^2 \rceil}$. □

We note that [12, Theorem 4] is obtained as a corollary of Theorem 4.5 by setting $w = 2$. Also, [17, Theorem 3.14] is obtained as a corollary by setting $N = q + 1$.

4.2 Nonconstructive Existence Results

The probabilistic method can be used to provide nonconstructive existence results for several of the types of codes considered in this paper. This approach was first used for PHF by Mehlhorn (see, for example, [14]), and a recent improvement using the Lovász Local Lemma can be found in [3]. A uniform approach to probabilistic bounds for PHF, SHF, and CFF, as well as a summary of known results, was given in [16].

The probabilistic method was used to prove an existence result for TA codes; the following theorem was proved in [6].

Theorem 4.6 *[6] There exists an (N, n, q) w-TA code, where $q = 2w^2$ and $N = 4w^2 \log n$.*

An examination of the proof of Theorem 4.6 shows that the code produced is in fact a $(w, 1 - 1/w)$-CFC. However, Theorem 4.6 does not compare favourably with the explicit construction presented in Theorem 4.5 — surprisingly, the explicit construction using Reed-Solomon codes yields better TA codes than the probabilistic method of Theorem 4.6. This in fact refutes the assertion made in [6], where it is claimed that no explicit constructions are known that are as efficient as Theorem 4.6. However, in the case of 2-IPP codes, the probabistic method was used in [12] to prove a result that is an improvement over Theorem 4.5.

Theorem 4.7 *[12] For every N and q, an (N, n, q) 2-IPP code exists in which $n \geq c(q/4)^{\frac{N}{3}}$, where $c = (27/32)^{1/3}$.*

5 Open Problems

Here is a list of interesting open problems and work points.

1. Do there exist w-IPP codes with $w < q < \lfloor (w+2)^2/4 \rfloor$?

2. Can we construct "interesting" w-TA codes with $q < w^2$? (Note: If $q < w^2$ in Theorem 4.5, then the code obtained has $n = q$ and hence is not interesting.)

3. It is not hard to construct examples of $(w, 1-1/w)$-CFC that do not have minimum distance $d > N(1-1/w^2)$. Therefore the converse of Theorem 4.3 does not hold. On the other hand, we do not know if the converse of Theorem 4.4 is true. Therefore we ask if there exist w-TA codes that are not $(w, 1-1/w)$-CFC.

4. Is there a "tight" characterization of w-IPP codes for $w \geq 3$. For example, if $w = 3$, we only know that 6-PHF $\Rightarrow$ 3-IPP $\Rightarrow$ (4-PHF + $(3,3)$-SHF).

5. Can we find nice explicit constructions for 2-SFPC and 2-FPC codes for arbitrary q, for example by modifying constructions in [12, 3, 4]?

6. Scalability: Can codes be embedded in larger codes with the same properties, by increasing N and n simultaneuosly? (Some constructions for embedding w-TA codes are given in [17, §4].)

7. w-TA codes provide traceability in linear time (i.e., in time $O(n)$.) In general, given a w-IPP code, traceability can be done in time $O(\binom{n}{w})$. Can this be improved, perhaps for certain subclasses of w-IPP codes?

Acknowledgement

D. R. Stinson's research was supported by NSERC grants IRC # 216431-96 and RGPIN # 203114-98.

References

[1] N. Alon and J. Spencer. *The Probabilistic Method*, Wiley, 1992.

[2] S. R. Blackburn. Combinatorics and threshold cryptography, in "Combinatorial Designs and their Applications", Chapman and Hall/CRC Research Notes in Mathematics, vol. 403, 1999, 49–70.

[3] S. R. Blackburn. Perfect hash families: probabilistic methods and explicit constructions, *Journal of Combinatorial Theory A*, to appear.

[4] S. R. Blackburn and P. R. Wild. Optimal linear perfect hash families, *Journal of Combinatorial Theory A* **83**(1998), 233-250.

[5] D. Boneh and J. Shaw. Collusion-secure fingerprinting for digital data, *IEEE Transactions on Information Theory* **44** (1998), 1897–1905.

[6] B. Chor, A. Fiat and M. Naor. Tracing traitors, in "Advances in Cryptology – Crypto '94", *Lecture Notes in Computer Science* **839** (1994), 480–491.

[7] Z. J. Czech, G. Havas and B. S. Majewski. Perfect hashing, *Theoretical Computer Science* **182** (1997), 1–143.

[8] Y. Desmedt, R. Safavi-Naini, H. Wang, C. Charnes and J. Pieprzyk. Broadcast anti-jamming systems, *ICON '99*, to appear.

[9] P. Erdös, P. Frankl and Z. Füredi. Families of finite sets in which no set is covered by the union of r others, *Israel Journal of Mathematics* **51**(1985), 79–89.

[10] A. Fiat and T. Tassa. Dynamic traitor tracing, in "Advances in Cryptology – Crypto '99", *Lecture Notes in Computer Science* **1666** (1999), 354–371.

[11] E. Gafni, J. Staddon and Y. L. Yin. Efficient methods for integrating traceability and broadcast encryption, in "Advances in Cryptology – Crypto '99", *Lecture Notes in Computer Science* **1666** (1999), 372–387.

[12] H. D. L. Hollmann, J. H. van Lint, J-P. Linnartz and L. M. G. M. Tolhuizen. On codes with the identifiable parent property, *Journal of Combinatorial Theory A* **82** (1998), 121–133.

[13] R. Kumar, S. Rajagopalan and A. Sahai. Coding constructions for blacklisting problems without computational assumptions, in "Advances in Cryptology – Crypto '99", *Lecture Notes in Computer Science* **1666** (1999), 609–623.

[14] K. Mehlhorn. *Data Structures and Algorithms 1: Sorting and Searching*, Springer-Verlag, 1984.

[15] R. Safavi-Naini and H. Wang. New results on multi-receiver authentication codes, in "Advances in Cryptology – Eurocrypt '98", *Lecture Notes in Computer Science* **1438** (1998), 527–541.

[16] D. R. Stinson, Tran van Trung and R. Wei. Secure frameproof codes, key distribution patterns, group testing algorithms and related structures, *Journal of Statistical Planning and Inference*, to appear.

[17] D. R. Stinson and R. Wei. Combinatorial properties and constructions of traceability schemes and frameproof codes, *SIAM Journal on Discrete Mathematics* **11** (1998), 41–53.