FRAMEPROOF CODES*

SIMON R. BLACKBURN[†]

Abstract. Frameproof codes were first introduced by Boneh and Shaw in the context of digital fingerprinting. Variants of these codes have been studied by several authors, and several similar definitions of frameproof codes exist in the literature. The paper considers frameproof codes from a combinatorial point of view, where we define frameproof codes as follows.

Let F be a (finite) set, and let $P \subseteq F^{\ell}$ be a set of words of length ℓ over the alphabet F. The set of descendants of P, desc(P), is the set of all words $x \in F^{\ell}$ such that for all $i \in \{1, 2, \ldots, \ell\}$, the ith component of x agrees with the *i*th component of some member of P. Let c be an integer such that $c \geq 2$. A c-frameproof code is a subset $C \subseteq F^{\ell}$ such that for all $P \subseteq C$ with $|P| \leq c$, we have that desc $(P) \cap C = P$.

The paper considers the following question: What is the largest cardinality n of a c-frameproof code of length ℓ , over an alphabet of size q? The paper concentrates on the case when q is large. The paper shows that $n = \ell(q-1)$ in the case when $2 \le \ell \le c$ and shows that if c = 2, then n is approximately $tq^{\lceil \ell/2 \rceil}$, where t = 1 when ℓ is odd and t = 2 if ℓ is even. The paper establishes improved upper bounds on n by applying techniques from extremal set theory (namely, a generalization of the Erdős–Ko–Rado theorem).

Key words. frameproof codes, digital fingerprinting, watermarking

AMS subject classification. 68R05

DOI. 10.1137/S0895480101384633

1. Introduction. Frameproof codes were first introduced by Boneh and Shaw [3] in the context of digital fingerprinting. There is more than one definition of frameproof codes in the literature; we use the following version.

Let F be a (finite) set of cardinality q and let ℓ be a positive integer. For a q-ary codeword $x \in F^{\ell}$ and an integer $i \in \{1, 2, \ldots, \ell\}$ we write x_i for the *i*th component of x. Let $P \subseteq F^{\ell}$ be a set of codewords of length ℓ . The set of descendants of P, desc(P), is the set of all words $x \in F^{\ell}$ such that for all $i \in \{1, 2, \ldots, \ell\}$, there exists $y \in P$ such that $x_i = y_i$. Let c be an integer such that $c \ge 2$. A c-frameproof code is a subset $C \subseteq F^{\ell}$ such that for all $P \subseteq C$ with $|P| \le c$, we have that desc $(P) \cap C = P$.

Boneh and Shaw use a different definition of descendant. The definition for frameproof codes we use is explicitly given by Fiat and Tassa [9], who credit Chor, Fiat, and Naor [4] with its first use. See Stinson and Wei [13] and Staddon, Stinson, and Wei [12] for constructions of binary frameproof codes and for a discussion of the relationship between frameproof codes and such concepts as traceability codes and codes with the identifiable parent property.

Inspired by an open question of Staddon, Stinson, and Wei [12, Section 5], we ask the following: What is the largest cardinality $M_{\ell,c}(q)$ of a *q*-ary *c*-frameproof code of length ℓ ? Let ℓ and *c* be fixed. We are interested in how $M_{c,\ell}(q)$ behaves as a function of *q*.

When $\ell \leq c$, we give a simple argument (Corollary 3) to show that $M_{c,\ell}(q) = \ell(q-1)$ for $q \geq 2$. The more interesting and difficult case is when $\ell > c$. As a first approximation, previous results (see Theorem 1 and Construction 2 below) imply that

^{*}Received by the editors February 6, 2001; accepted for publication (in revised form) February 6, 2003; published electronically June 25, 2003.

http://www.siam.org/journals/sidma/16-3/38463.html

[†]Department of Mathematics, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK (s.blackburn@rhul.ac.uk).

 $M_{c,\ell}(q) = \Theta(q^{\lceil \ell/c \rceil})$, where the constants hidden by the notation may depend on ℓ and c. This suggests that we examine the behavior of the ratio $R_{c,\ell}(q)$ defined by $R_{c,\ell}(q) = M_{c\ell}(q)/q^{\lceil \ell/c \rceil}$. Define t to be the unique integer such that $1 \leq t \leq c$ and $t = \ell \mod c$. Again, it follows from known results that $\overline{\lim}_{q\to\infty} R_{c,\ell}(q) \leq \max\{1,t\}$ and that $\underline{\lim}_{q\to\infty} \geq 1$. When $\ell = 1 \mod c$, these results imply (Corollary 5) that $\lim_{q\to\infty} R_{c,\ell}(q)$ exists and is equal to 1.

One case not covered by the above is the case c = 2 and ℓ even, where the above results show that $1 + o(1) \leq R_{c,\ell}(q) \leq 2 + o(1)$. In section 4, we give a construction that matches the upper bound, thus establishing that $\lim_{q\to\infty} R_{c,\ell}(q) = 2$ in this case.

In sections 5 and 6, we turn to improving the upper bound. Defining t as above, we show that $R_{c,\ell}(q) \leq \ell/(\ell - (t-1)\lceil \ell/c \rceil)$ by relating the problem of providing an upper bound to a problem in extremal set theory. In the two cases when t = 1 and t = c, this bound is essentially the same as the upper bound of t + o(1) given by Theorem 1, but for any other values of t and c it gives an improvement. Indeed, when c is fixed and ℓ is large, then our new upper bound is approximately c/(c - t + 1) which is generally much less than t.

In general there is still a gap between the upper bounds we have given for $R_{c,\ell}(q)$ and the lower bounds that follow from known large q-ary c-frameproof codes of length ℓ . In section 7 we close this gap in one case: when $\ell = 5$ and c = 3. By constructing a code of size $(5/3)q^2 + O(q)$, we show that our upper bound on $R_{c,\ell}(q)$ of 5/3 + o(1)is tight when $q \to \infty$ by establishing that $R_{c,\ell}(q) = 5/3 + o(1)$.

Note that any set of length 1 vectors is a c-frameproof code for any c; thus the length 1 case is trivial. For the remainder of the paper we consider codes of length ℓ , where $\ell \geq 2$.

The paper is organized as follows. Section 2 proves an upper bound on the size of a *c*-frameproof code. This bound is a slight modification of the bound given in Staddon, Stinson, and Wei [12, Theorem 3.7]. Section 3 contains two constructions of *q*-ary *c*-frameproof codes of length ℓ (one of these constructions has been given before, in Cohen and Encheva [5, Proposition 1]). Section 4 contains a third, more complicated, construction of 2-frameproof codes. The constructions of sections 3 and 4 show that the leading term of the upper bound has the correct order of magnitude; moreover, the leading coefficient of the upper bound is tight when c = 2, when $\ell \leq c$, or when $\ell = 1 \mod c$. Section 5 improves the bound of section 2 by relating the problem to a question in the theory of intersecting systems of finite sets. This set theoretic question is investigated further in section 6. Section 7 constructs a family of 3-frameproof codes of length 5 to show that the improved upper bound given in sections 5 and 6 is tight in this case. Finally, the paper ends with a brief discussion of open problems.

2. An upper bound.

THEOREM 1. Let ℓ , q, and c be positive integers such that $c \ge 2$ and $\ell \ge 2$. Let C be a q-ary c-frameproof code of length ℓ with cardinality n greater than q. Define the integer $r \in \{0, 1, \ldots, c-1\}$ to be the remainder of ℓ on division by c. Then

(1)
$$n \le \max\left\{q^{\lceil \ell/c \rceil}, r\left(q^{\lceil \ell/c \rceil} - 1\right) + (c - r)\left(q^{\lfloor \ell/c \rfloor} - 1\right)\right\}.$$

We remark that for almost all parameter sets, the second term on the right-hand side of (1) is the largest.

FRAMEPROOF CODES

Proof. Let C be a q-ary length ℓ c-frameproof code of cardinality n. We show that the bound (1) holds. For any subset $S \subseteq \{1, 2, \ldots, \ell\}$, define U_S by

 $U_S = \{x \in C : \text{ there exists no } y \in C \setminus \{x\} \text{ such that } x_i = y_i \text{ for all } i \in S\}.$

Note that $|U_S| \leq q^{|S|}$, since every codeword $x \in U_S$ is uniquely identified by the subword $(x_i : i \in S)$. Moreover, if $n > q^{|S|}$ then $|U_S| \leq q^{|S|} - 1$, since at least one choice of the subword $(x_i : i \in S)$ must correspond to two or more codewords in C.

Let $S_1, S_2, \ldots, S_c \subseteq \{1, 2, \ldots, \ell\}$ be disjoint subsets, where $|S_j| = \lceil \ell/c \rceil$ whenever $1 \leq j \leq r$ and $|S_j| = \lfloor \ell/c \rfloor$ whenever $r+1 \leq j \leq c$. So $\cup_{j=1}^c S_j = \{1, 2, \ldots, \ell\}$. The bound of the theorem follows if we can show that $C = \cup_{j=1}^c U_{S_j}$.

Suppose, for a contradiction, that $x \in C \setminus \bigcup_{j=1}^{c} U_{S_j}$. So there exist $x^1, x^2, \ldots, x^c \in C \setminus \{x\}$ such that x^j and x agree in their *i*th components for all $i \in S_j$. But then $x \in \operatorname{desc}(\{x^1, x^2, \ldots, x^c\})$, which contradicts the *c*-frameproof property of *C*. This contradiction shows that $C = \bigcup_{j=1}^{c} U_{S_j}$, as required. \Box

COROLLARY 2. A q-ary c-frameproof code of length ℓ contains at most

$$tq^{\lceil \ell/c \rceil} + O(q^{\lceil \ell/c \rceil - 1})$$

codewords, where t is the unique integer such that $t \in \{1, 2, \ldots, c\}$ and $t = \ell \mod c$.

3. Two constructions. This section presents two constructions of frameproof codes; the second of these constructions is given in Cohen and Encheva [5, Proposition 1].

CONSTRUCTION 1. Let $F = \{0, 1, ..., q - 1\}$. The set C of all words of length ℓ and weight exactly 1 (i.e., the elements of F^{ℓ} with exactly one nonzero component) forms a c-frameproof code of cardinality $\ell(q - 1)$.

Proof. Let $x \in C$ be a weight 1 vector, and suppose its *i*th component is nonzero. Now, any set $P \subseteq C$ such that $x \in \operatorname{desc}(P)$ must contain a codeword y such that $y_i = x_i$. But since a codeword of weight 1 is uniquely determined by its nonzero component, we must have that x = y. Hence C is c-frameproof for any c. \Box

Theorem 1 and Construction 1 combine to show the following result.

COROLLARY 3. Let q, ℓ , and c be positive integers such that $q \ge 2$ and $2 \le \ell \le c$. Then the largest q-ary length ℓ c-frameproof code has cardinality $\ell(q-1)$.

CONSTRUCTION 2. Let integers ℓ and c be such that $\ell \geq 2$ and $c \geq 2$. Let q be a prime power such that $q \geq \ell$. Let F be the finite field of cardinality q and let $\alpha_1, \alpha_2, \ldots, \alpha_\ell \in F$ be distinct. Define a length ℓ code C over F by

$$C = \{ (f(\alpha_1), f(\alpha_2), \dots, f(\alpha_\ell)) : f \in F[X] \text{ and } \deg f < \lceil \ell/c \rceil \}.$$

Then C is a c-frameproof code of cardinality $q^{\lceil \ell/c \rceil}$.

We remark that the restriction $q \ge \ell$ may be weakened to $q+1 \ge \ell$ by also allowing a polynomial f to be evaluated at a "point at infinity": $f(\infty)$ is defined to be the coefficient of $X^{\lceil \ell/c \rceil - 1}$ in f.

Proof. There are $q^{\lceil \ell/c \rceil}$ choices for a polynomial f of degree less than $\lceil \ell/c \rceil$ as there are q choices for each of its coefficients.

If $x, y \in C$ agree in $\lceil \ell/c \rceil$ positions, then x = y (since we may recover the polynomial associated with a codeword by interpolation by considering just the positions where x and y agree). In particular, each distinct choice for the polynomial f gives rise to a distinct codeword, since f is determined by specifying $f(\alpha)$ at $\lceil \ell/c \rceil$ points α . Hence $|C| = q^{\lceil \ell/c \rceil}$. Now, let $x \in C \cap \operatorname{desc}(P)$, where $P \subseteq C$ has cardinality at

most c. Each component of x must agree with the corresponding component of one of the codewords in P, and so there is a codeword $y \in P$ that agrees with x in at least $\lceil \ell/c \rceil$ positions. But then $x = y \in P$, and so the code is c-frameproof. \Box

Corollary 2 and Construction 2 combine to show the following two results.

COROLLARY 4. Let ℓ and c be fixed integers such that $\ell \geq 2$ and $c \geq 2$. Let $M_{c,\ell}(q)$ be the largest cardinality of a q-ary c-frameproof code of length ℓ . Then

$$\lim_{q \to \infty} \log_q M_{c,\ell}(q) = \lceil \ell/c \rceil.$$

Proof. We have that $\log_q M_{c,\ell}(q) \leq \lceil \ell/c \rceil + o(1)$ by Corollary 2.

For a given value of q, let q' be the largest prime power such that $q' \leq q$. By the prime number theorem, q'/q = 1 - o(1). By Construction 2, we have that $\log_{q'} M_{c,\ell}(q') \geq \lceil \ell/c \rceil$ whenever q' is sufficiently large. Hence

$$\log_q M_{c,\ell}(q) \ge \log_q M_{c,\ell}(q') \ge \log_{q'} M_{c,\ell}(q') - o(1) \ge \lceil \ell/c \rceil - o(1).$$

These bounds on $\log_q M_{c,\ell}(q)$ imply that $\lim_{q\to\infty} \log_q M_{c,\ell}(q)$ exists and is equal to $\lceil \ell/c \rceil$, as required. \Box

The proof of the following corollary is similar to the proof of Corollary 4.

COROLLARY 5. Let ℓ and c be fixed integers such that $\ell \geq 2$, $c \geq 2$, and $\ell = 1 \mod c$. Let $M_{c,\ell}(q)$ be defined as in Corollary 4. Then

$$\lim_{q \to \infty} M_{c,\ell}(q)/q^{\lceil \ell/c \rceil} = 1.$$

4. 2-frameproof codes of even length. We aim to construct a family of 2-frameproof codes of length ℓ , where ℓ is even. This construction, when combined with Construction 2, will show that the leading term of the upper bound given in Theorem 1 is tight in the case when c = 2.

We define two subcodes as part of our final construction. Let ℓ be an even integer such that $\ell \geq 4$. Let m be a prime power such that $m \geq \ell + 1$ and set $q = m^2 + 1$. Let \mathbb{F}_m be the finite field of order m, and define F to be the disjoint union $F = \{\infty\} \cup (\mathbb{F}_m)^2$. Let $\beta_0, \beta_1, \alpha_1, \alpha_2, \ldots, \alpha_{\ell-1}$ be distinct elements of \mathbb{F}_m . For polynomials $f, g \in \mathbb{F}_m[X]$, we write $(f, g)(\alpha_i)$ for the element $(f(\alpha_i), g(\alpha_i)) \in F$. Define $C_1 \subseteq F^{\ell}$ by

(2)
$$C_1 = \{ (\infty, (f,g)(\alpha_1), (f,g)(\alpha_2), \dots, (f,g)(\alpha_{\ell-1})) \},\$$

where $f, g \in \mathbb{F}_m[X]$ are such that deg $f = (\ell/2) - 1$ and deg $g \leq (\ell/2) - 1$. Define $C_2 \subseteq F^{\ell}$ by

(3)
$$C_2 = \{((t(\beta_0), t(\beta_1)), (s, t)(\alpha_1), (s, t)(\alpha_2), \dots, (s, t)(\alpha_{\ell-1}))\},\$$

where $s, t \in \mathbb{F}_m[X]$ are such that deg $s \leq (\ell/2) - 2$ and deg $t \leq (\ell/2)$.

CONSTRUCTION 3. Let ℓ be an even integer such that $\ell \geq 4$. Let m be a prime power such that $m \geq \ell + 1$ and set $q = m^2 + 1$. Define C_1 and C_2 as above. Then the code C defined by $C = C_1 \cup C_2$ is a 2-frameproof code of cardinality $2(q-1)^{\ell/2}(1-1/(2\sqrt{q-1}))$.

Proof. By considering their first components, it is clear that C_1 and C_2 are disjoint. A polynomial of degree at most $(\ell/2) - 1$ is determined by its values at $\ell/2$ distinct points, and hence the polynomials f and g in (2) are uniquely determined by a codeword $x \in C_1$. There are $m^{\ell/2} - m^{(\ell/2)-1}$ choices for f and there are

502

 $m^{\ell/2}$ choices for g, and so $|C_1| = (m^2)^{\ell/2}(1-1/m)$. The polynomial s in (3) is determined by $(\ell/2)$ of the final $\ell - 1$ components of a codeword $x \in C_2$. Similarly, the polynomial t is determined by $(\ell/2) + 1$ of these components. Hence $|C_2|$ is equal to the number of choices for s and t and so $|C_2| = m^{(\ell/2)-1}m^{(\ell/2)+1} = (m^2)^{\ell/2}$. Summing our expressions for $|C_1|$ and $|C_2|$ and using the fact that $m = \sqrt{q-1}$ shows that $|C| = 2(q-1)^{\ell/2}(1-1/(2\sqrt{q-1}))$, as required.

It remains to show that C is a 2-frameproof code. To this end, we claim that codewords $x \in C_1$ and $y \in C_2$ can agree in at most $(\ell/2) - 1$ components. The first components of x and y are never equal. If $\ell/2$ of the remaining positions agree, then the definitions of C_1 and C_2 imply that a polynomial f of degree exactly $(\ell/2) - 1$ and a polynomial s of degree at most $(\ell/2) - 2$ agree at $\ell/2$ points. This contradiction establishes our claim.

Let $P \subseteq C$ be such that |P| = 2. Let $x \in \operatorname{desc}(P) \cap C$. We must show that $x \in P$. Suppose that $x \in C_1$. Excluding the first coordinate, there are $\ell - 1$ coordinates, and so x must agree with some member $y \in P$ in $\lceil \ell/2 \rceil = \ell/2$ positions other than the first. Since x and y agree in more than $(\ell/2) - 1$ positions, we must have that $y \in C_1$. But any $\ell/2$ of the last $\ell - 1$ components determine a codeword in C_1 , and so x = y. Hence $x = y \in P$, as required.

Now suppose that $x \in C_2$. Let $y \in P$ be such that $x_1 = y_1$ (and so $y \in C_2$). If x and y agree on $(\ell/2) - 1$ or more of the last $\ell - 1$ components, then the components on which x and y agree include $(\ell/2) - 1$ values of s and $(\ell/2) + 1$ values of t, and so x = y. Thus $x = y \in P$ in this case. Now suppose that x and y agree on less than $(\ell/2) - 1$ of the last $\ell - 1$ components. If we define z to be the element of P not equal to y, we have that x and z agree in at least $(\ell/2) + 1$ components. This implies that $z \in C_2$, and since the components on which x and z agree include at least $\ell/2$ values of s and $(\ell/2) + 1$ values of t, we have that x = z. Hence $x = z \in P$ in this case also, and so C is a 2-frameproof code. \Box

COROLLARY 6. In the notation of Corollary 4,

$$\lim_{q \to \infty} M_{2,\ell}(q)/q^{\lceil \ell/2 \rceil} = 1 \text{ when } \ell \text{ is odd,}$$
$$\lim_{q \to \infty} M_{2,\ell}(q)/q^{\lceil \ell/2 \rceil} = 2 \text{ when } \ell \text{ is even.}$$

5. An improved upper bound. Given Corollaries 3 and 6, it might be tempting to conjecture that the leading term of Theorem 1 is always tight. However, this is not the case. This section reduces the problem of providing an improved upper bound to a problem in extremal set theory. This latter problem will be considered in section 6.

Let ℓ and k be fixed integers, where $1 \leq k \leq \ell$. Let D be a set, and let

$$(V_S \subseteq D : S \subseteq \{1, 2, \dots, \ell\}, |S| = k)$$

be a family of subsets of D indexed by the subsets of $\{1, 2, \ldots, \ell\}$ of cardinality k. We say that this family is a $(k, \ell; b, t)$ -frameproof code set system (FPCSS) if $|V_S| \leq b$ for all subsets S of $\{1, 2, \ldots, \ell\}$ of cardinality k, and if

$$(4) V_{S_1} \cup V_{S_2} \cup \dots \cup V_{s_t} = D$$

whenever S_1, S_2, \ldots, S_t are pairwise disjoint subsets of $\{1, 2, \ldots, \ell\}$ of cardinality k. We define the *size* of a $(k, \ell; b, t)$ -FPSS to be |D|. We aim to show (see Lemma 7) that a frameproof code gives rise to an FPCSS of comparable size. If we can determine the largest size of an FPCSS, then this will provide an upper bound on the size of a frameproof code.

LEMMA 7. Let q, c, and ℓ be positive integers, and suppose that $\ell > c$. Let C be a q-ary c-frameproof code of length ℓ containing n codewords. Let $t \in \{1, 2, \ldots, c\}$ be such that $t = \ell \mod c$. Let $k = \lceil \ell/c \rceil$. Then there exists a $(k, \ell; q^k, t)$ -FPCSS of size at least

$$n - \binom{\ell}{k-1} q^{k-1}.$$

Proof. As in section 2, for any $S \subseteq \{1, 2, \ldots, \ell\}$ we define U_S to be the set of codewords $x \in C$ which are uniquely determined by the ordered subset $(x_i : i \in S)$ of their components. Just as in the proof of Theorem 1, we may show that

$$C = U_{S_1} \cup U_{S_2} \cup \cdots \cup U_{S_c},$$

whenever S_1, S_2, \ldots, S_c are subsets of $\{1, 2, \ldots, \ell\}$ with the property that $S_1 \cup S_2 \cup \cdots \cup S_c = \{1, 2, \ldots, \ell\}$.

We define an FPCSS as follows. Let

$$D = C \setminus \left(\bigcup_{S} U_{S}\right),$$

where S runs through all subsets of $\{1, 2, \ldots, \ell\}$ of cardinality k - 1. We observed in the proof of Theorem 1 that $|U_S| \leq q^{|S|}$, and so

$$|D| \ge n - \binom{\ell}{k-1} q^{k-1}.$$

For any subset $S \subseteq \{1, 2, \dots, \ell\}$ such that |S| = k, we define

$$V_S = U_S \cap D$$

Clearly, $|V_S| \leq |U_S| \leq q^k$.

It remains to show that the subsets V_S do indeed form a $(k, \ell; q^k, t)$ -FPCSS. Let S_1, S_2, \ldots, S_t be a set of pairwise disjoint subsets of $\{1, 2, \ldots, \ell\}$ of cardinality k. We need to show that $V_{S_1} \cup V_{S_2} \cup \cdots \cup V_{S_t} = D$. The number of elements of $\{1, 2, \ldots, \ell\}$ which are not contained in $S_1 \cup S_2 \cup \cdots \cup S_t$ is $\ell - tk = (c - t)(k - 1)$. Hence there exist subsets $S_{t+1}, S_{t+2}, \ldots, S_c$ of cardinality k - 1 such that

$$S_1 \cup S_2 \cup \cdots \cup S_c = \{1, 2, \dots, \ell\}.$$

By our definition of D, we have that $U_{S_i} \cap D = \emptyset$ whenever $i \ge t+1$. Hence

$$V_{S_1} \cup V_{S_2} \cup \dots \cup V_{S_t} = (U_{S_1} \cup U_{S_2} \cup \dots \cup U_{S_t}) \cap D$$
$$= (U_{S_1} \cup U_{S_2} \cup \dots \cup U_{S_c}) \cap D$$
$$= C \cap D$$
$$= D.$$

Thus our sets form an FPCSS as claimed, and so the lemma follows. \Box

We now introduce the problem in extremal set theory that we will be concerned with. We say that a family S of subsets of a set is *t*-colliding if S does not contain *t* pairwise disjoint subsets.

Let t, k, and ℓ be positive integers such that $1 \leq k \leq \ell$. We define $m(t, k, \ell)$ to be the maximum number of subsets in a *t*-colliding family S of subsets of $\{1, 2, \ldots, \ell\}$, where |S| = k for all $S \in S$. Note that $m(t, k, \ell) = \binom{\ell}{k}$ when $tk > \ell$, and $m(t, k, \ell) < \binom{\ell}{k}$ otherwise.

THEOREM 8. Let t, k, ℓ , and b be positive integers such that $tk \leq \ell$. Then a $(k, \ell; b, t)$ -FPCSS has size at most

$$\left(\frac{1}{1-m(t,k,\ell)/\binom{\ell}{k}}\right)b.$$

We remark that when $tk > \ell$, the condition (4) becomes trivial and so there is no bound on the size of a $(k, \ell; b, t)$ -FPCSS.

Proof. Let D be a set, and let (V_S) be a collection of subsets of D that forms a $(k, \ell; b, t)$ -FPCSS. We prove our upper bound on |D| by counting, in two ways, the elements of the set

(5)
$$K = \{(x, S) : x \in V_S\},\$$

where $S \subseteq \{1, 2, \dots, \ell\}$ is such that |S| = k, and where $x \in D$.

There are $\binom{\ell}{k}$ choices for the subset S. Once S is chosen, there are at most b choices for x since $|V_S| \leq b$ by the definition of an FPCSS. Hence $|K| \leq \binom{\ell}{k}b$.

We claim that an element $x \in D$ is contained in V_S for at least $\binom{\ell}{k} - m(t, k, \ell)$ subsets S of cardinality k. Let S be defined by

$$\mathcal{S} = \{ S \subseteq \{1, 2, \dots, \ell\} : |S| = k \text{ and } x \notin V_S \}.$$

Now, S is t-colliding, for if there exist pairwise disjoint subsets $S_1, S_2, \ldots, S_t \in S$, then $x \notin V_{S_1} \cup V_{S_2} \cup \cdots \cup V_{S_t}$, which would contradict the FPCSS property (4). Since S is t-colliding, $|S| \leq m(t, k, \ell)$, and so our claim follows.

There are |D| choices for the element x in (5), and our claim implies that once x is fixed, there are at least $\binom{\ell}{k} - m(t,k,\ell)$ choices for S such that $(x,S) \in K$. Hence $|K| \ge |D| \binom{\ell}{k} - m(t,k,\ell)$. But now

$$|D|(\binom{\ell}{k} - m(t, k, \ell)) \le |K| \le \binom{\ell}{k}b,$$

and so the theorem follows.

The bound of Theorem 8 is tight, as the following example shows. Let t, k, and ℓ be positive integers, and suppose that $tk \leq \ell$. Let S be a *t*-colliding family of subsets of $\{1, 2, \ldots, \ell\}$ with the property that |S| = k for all $S \in S$, and suppose that S consists of $m(t, k, \ell)$ subsets. Define $D = \text{Sym}(\ell)$, the symmetric group on ℓ letters. For any subset $S \subseteq \{1, 2, \ldots, \ell\}$ such that |S| = k, define

$$V_S = \{ \pi \in D : \pi(S) \notin \mathcal{S} \}$$

Let S_1, S_2, \ldots, S_t be pairwise disjoint subsets of $\{1, 2, \ldots, \ell\}$ with $|S_i| = k$ for all $i \in \{1, 2, \ldots, t\}$. Let $\pi \in D$ and suppose that $\pi \notin V_{S_1} \cup V_{S_2} \cup \cdots \cup V_{S_t}$. Then $\pi(S_i) \in \mathcal{S}$ for all $i \in \{1, 2, \ldots, t\}$ by the definition of V_S . But this implies that $\pi(S_1), \pi(S_2), \ldots, \pi(S_t)$ form a set of t pairwise disjoint subsets in \mathcal{S} , contradicting the fact that \mathcal{S} is t-colliding. Hence $\pi \in V_{S_1} \cup V_{S_2} \cup \cdots \cup V_{S_t}$ for all $\pi \in D$, and condition (4) follows. It is easy to see that $|D| = \ell!$ and that the sets V_S all have cardinality $b = (\binom{\ell}{k} - m(t,k,\ell))k!(\ell-k)!$. Hence D is a $(k,\ell;b,t)$ -FPCSS that meets the bound of Theorem 8, as required.

COROLLARY 9. Let c and ℓ be integers, and suppose that $c \ge 2$ and $\ell \ge 2$. Let $t \in \{1, 2, \ldots, c\}$ be such that $t = \ell \mod c$. Let C be a q-ary c-frameproof code of length ℓ . As $q \to \infty$ with c and ℓ fixed, we have that

$$|C| \le \kappa q^{\lceil \ell/c \rceil} + O(q^{\lceil \ell/c \rceil - 1}),$$

where κ is the constant defined by

$$\kappa = \frac{1}{1 - m(t, \lceil \ell/c \rceil, \ell) / \binom{\ell}{\lceil \ell/c \rceil}}.$$

Proof. The corollary follows by Lemma 7 and Theorem 8 after observing that $t\lceil \ell/c\rceil \leq \ell$. \Box

6. Intersecting set systems. Recall from the previous section that a family of subsets is *t*-colliding if it does not contain a set of *t* pairwise disjoint subsets. Let *t*, *k*, and ℓ be positive integers such that $tk \leq \ell$. Define, as before, $m(t, k, \ell)$ to be the maximum size of a *t*-colliding family S of subsets of $\{1, 2, \ldots, \ell\}$ such that |S| = k for all $S \in S$. This section proves an upper bound on $m(t, k, \ell)$.

Note that the case when t = 1 is trivial: no nonempty family of subsets can be 1-colliding, and so $m(1, k, \ell) = 0$ in this case. We will therefore assume that $t \ge 2$.

The family \mathcal{M} defined by

$$\mathcal{M} = \{ S \subseteq \{1, 2, \dots, \ell\} : |S| = k \text{ and } S \cap \{1, 2, \dots, t-1\} \neq \emptyset \}$$

is clearly t-colliding, and $|\mathcal{M}| = {\binom{\ell}{k}} - {\binom{\ell-(t-1)}{k}}$. This family provides a lower bound on $m(t, k, \ell)$, which we would expect to be realistic. Indeed, much of the literature on this problem has been concerned with showing that \mathcal{M} is optimal (in the sense that $m(t, k, \ell) = |\mathcal{M}|$) when certain conditions on t, k, and ℓ are met. The famous theorem of Erdős, Ko, and Rado [8] (see Anderson [1]) asserts in our notation that $m(2, k, \ell) = {\binom{\ell-1}{k-1}}$, and so \mathcal{M} is optimal in the case when t = 2. Erdős [7] was the first to consider the problem when t > 2; he proves that there exists a constant κ depending only on k such that \mathcal{M} is optimal whenever $\ell > \kappa t$. Bollobás, Daykin, and Erdős [2] show that $\ell > 2k^3t$ will suffice. In Deza and Frankl [6, section 4], a result of Frankl is mentioned that shows that \mathcal{M} is optimal whenever $\ell > \kappa' k t^2$ for some constant κ' . Deza and Frankl conjecture that \mathcal{M} is optimal whenever $\ell > \kappa''kt$ for some constant κ'' .

Rather than proving that $m(t, k, \ell) = \binom{\ell}{k} - \binom{\ell - (t-1)}{k}$ for certain values of t, k, and ℓ , we would like an upper bound on $m(t, k, \ell)$ that holds for any values of t, k, and ℓ . Such a bound is given in Theorem 11 below. This bound is inspired by Katona's proof [11] of the Erdős–Ko–Rado theorem and is a special case of a bound of Gronau [10]; we include a proof here for the sake of completeness.

Before proving Gronau's bound, we will first consider a simpler situation. Let \mathbb{Z}_{ℓ} denote the integers modulo ℓ . For $a \in \mathbb{Z}_{\ell}$, define $T_{\ell}(a) \subseteq \mathbb{Z}_{\ell}$ by

$$T_{\ell}(a) = \{a, a+1, a+2, \dots, a+(k-1)\}.$$

Write $\mathcal{T} = \{T_{\ell}(a) : a \in \mathbb{Z}_{\ell}\}.$

506

LEMMA 10. Let t, k, and ℓ be positive integers such that $\ell \geq tk$. Define the sets $T_{\ell}(a)$ and the family \mathcal{T} as above. Suppose that \mathcal{S} is contained in \mathcal{T} and is t-colliding. Then $|\mathcal{S}| \leq (t-1)k$.

We remark that the family $S = \{T_{\ell}(a) : 0 \le a \le (t-1)k-1\}$ is t-colliding and meets the bound of Lemma 10.

Proof. We prove the lemma by induction on ℓ . Suppose that $\ell = tk$. In this case, we may partition \mathcal{T} into parts $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_k$, where

$$\mathcal{T}_i = \{T_\ell(i), T_\ell(i+k), T_\ell(i+2k), \dots, T_\ell(i+(t-1)k)\}.$$

Since \mathcal{T}_i consists of t pairwise disjoint sets, \mathcal{T}_i is not contained in S and so $|\mathcal{T}_i \cap S| \leq 1$ t-1. Hence

$$\begin{split} |\mathcal{S}| &= |(\mathcal{T}_1 \cup \mathcal{T}_2 \cup \dots \cup \mathcal{T}_k) \cap \mathcal{S} \\ &= \sum_{i=1}^k |\mathcal{T}_i \cap \mathcal{S}| \\ &\leq (t-1)k, \end{split}$$

and so the lemma follows when $\ell = tk$.

Assume, as an inductive hypothesis, that $\ell > tk$ and the lemma holds for all smaller values of ℓ . Certainly $\mathcal{S} \neq \mathcal{T}$, and so there exists $c \in \mathbb{Z}_{\ell}$ such that $T_{\ell}(c) \notin \mathcal{S}$. We may define a family $\overline{\mathcal{S}}$ of subsets of $\mathbb{Z}_{\ell-1}$ by

$$\overline{\mathcal{S}} = \{ T_{\ell-1}(a) : a \in \{0, 1, \dots, c-1\}, T_{\ell}(a) \in S \}$$
$$\cup \{ T_{\ell-1}(a-1) : a \in \{c+1, c+2, \dots, \ell-1\}, T_{\ell}(a) \in S \}.$$

Clearly there is a one-to-one correspondence between the subsets in \mathcal{S} and the subsets in $\overline{\mathcal{S}}$, and so $|\mathcal{S}| = |\overline{\mathcal{S}}|$. Moreover, the cardinality of the intersection of a pair of subsets in $\overline{\mathcal{S}}$ is at least as great as the cardinality of the intersection of the corresponding pair of subsets in \mathcal{S} . Hence the fact that \mathcal{S} is t-colliding implies that $\overline{\mathcal{S}}$ is t-colliding. Our inductive hypothesis now implies that $|\overline{\mathcal{S}}| \leq (t-1)k$, and so $|\mathcal{S}| \leq (t-1)k$ as required. The lemma now follows by induction on ℓ . П

THEOREM 11. Let t, k, and ℓ be positive integers, where $tk \leq \ell$. Let S be a t-colliding family of subsets of $\{1, 2, \dots, \ell\}$, where |S| = k for all $S \in S$. Then

$$|\mathcal{S}| \le \binom{\ell}{k} \frac{(t-1)k}{\ell}.$$

So Theorem 11 states that $m(t,k,\ell) \leq {\ell \choose k} \frac{(t-1)k}{\ell}$. The bound of Theorem 11 is best possible when t = 1 (as the problem is trivial) and t = 2 (the *t*-colliding family \mathcal{M} defined near the start of this section provides the appropriate example). In the case when $tk = \ell$, the *t*-colliding family

$$\mathcal{N} = \{ S \subseteq \{1, 2, \dots, \ell\} : |S| = k \text{ and } 1 \notin S \}$$

contains $\binom{\ell}{k} \frac{(t-1)k}{\ell}$ sets. So Theorem 11 is also best possible in the case when $tk = \ell$. When t and k are fixed with $\ell \to \infty$, the upper bound on $m(t, k, \ell)$ provided by Theorem 11 has the form $(t-1)\ell^{k-1}/(k-1)! + O(\ell^{k-2})$. But the lower bound on $m(t, k, \ell)$ provided by the t-colliding family \mathcal{M} at the start of the section also has this form, as can be easily seen from the expression $|\mathcal{M}| = \sum_{i=1}^{t-1} {t-1 \choose i} {\ell-(t-1) \choose k-i}$. In

particular, the ratio between the upper and lower bounds on $m(t, k, \ell)$ tends to 1 as $\ell \to \infty$ with t and k fixed. So the upper bound of Theorem 11 is the right order of magnitude when ℓ is large.

Proof of Theorem 11. Define \mathcal{T} as above, and let Q be the set of pairs (α, S) , where $S \in \mathcal{S}$ and $\alpha : \{1, 2, \dots, \ell\} \to \mathbb{Z}_{\ell}$ is a bijection such that $\alpha(S) \in \mathcal{T}$. We will count the elements of Q in two ways.

There are $|\mathcal{S}|$ choices for $S \in \mathcal{S}$. Once S has been chosen, there are ℓ choices for $\alpha(S) \in \mathcal{T}$ and then $k!(\ell - k)!$ choices for a suitable bijection α . Hence

$$|Q| = \ell |\mathcal{S}| k! (\ell - k)!$$

We now count the elements of Q in a different way. There are $\ell!$ choices for α . Suppose now that α is fixed. The number of choices for S is $|\mathcal{X}|$, where $\mathcal{X} = \{S \in \mathcal{S} : \alpha(S) \in \mathcal{T}\}$. Now, \mathcal{X} is t-colliding because it is a subfamily of \mathcal{S} . Hence the corresponding subfamily $\alpha(\mathcal{X})$ of \mathcal{T} (where $\alpha(\mathcal{X}) = \{\alpha(S) : S \in \mathcal{X}\}$) is t-colliding. Hence $|\mathcal{X}| = |\alpha(\mathcal{X})| \leq (t-1)k$ by Lemma 10. So

$$|Q| \le \ell! (t-1)k_i$$

and therefore

$$\begin{split} |\mathcal{S}| &= |Q| / \left(k! (\ell - k)! \ell \right) \\ &\leq \ell! (t - 1)k / \left(k! (\ell - k)! \ell \right) \\ &= \binom{\ell}{k} \frac{(t - 1)k}{\ell}, \end{split}$$

as required. \Box

COROLLARY 12. Let c and ℓ be integers, and suppose that $c \ge 2$ and $\ell \ge 2$. Let $t \in \{1, 2, \ldots, c\}$ be such that $t = \ell \mod c$. Let C be a q-ary c-frameproof code of length ℓ . Then

$$|C| \leq \left(\frac{\ell}{\ell - (t-1)\lceil \ell/c\rceil}\right) q^{\lceil \ell/c\rceil} + O(q^{\lceil \ell/c\rceil-1}).$$

Proof. The corollary follows by combining Corollary 9 with Theorem 11. \Box

7. A 3-frameproof code of length 5. The first case where the upper bound of section 5 improves on the bound of section 2 is when we are considering q-ary 3-frameproof codes of length 5. The upper bound of section 5 shows that such a q-ary 3-frameproof code has cardinality at most $\frac{5}{3}q^2 + O(q)$. We will now show that the leading term of the bound is tight in this case by constructing a 3-frameproof code of length 5 of sufficiently large cardinality.

We define five sets X_1, X_2, X_3, X_4 , and X_5 of words of length 5 over the alphabet $\mathbb{F}_3 \cup \{\infty\}$ as follows:

| X_1 | = | {(| $\infty,$ | a, | a, | a, | a | $): a \in \mathbb{Z}_3$ |
|-------|---|----|-----------|-----------|-----------|-----------|----------|-------------------------|
| X_2 | = | {(| a, | $\infty,$ | a, | a + 1, | a+2 | $): a \in \mathbb{Z}_3$ |
| X_3 | = | {(| a, | a, | $\infty,$ | a + 2, | a+1 | $): a \in \mathbb{Z}_3$ |
| X_4 | = | {(| a, | a + 1, | a+2, | $\infty,$ | a | $): a \in \mathbb{Z}_3$ |
| X_5 | = | {(| a, | a + 2, | a + 1, | a, | ∞ |): $a \in \mathbb{Z}_3$ |

The sets X_i are clearly pairwise disjoint and have cardinality 3. Moreover, it is not difficult to check that a codeword in $X_1 \cup X_2 \cup X_3 \cup X_4 \cup X_5$ is uniquely determined by specifying two of its components.

Let m be a prime power such that $m \geq 4$. Let $\alpha_1, \alpha_2, \alpha_3$, and α_4 be distinct elements in \mathbb{F}_m . Define five sets Y_1, Y_2, Y_3, Y_4 , and Y_5 of words of length 5 over the alphabet $\mathbb{F}_m \cup \{\infty\}$ by

$$\begin{split} Y_1 &= \{ (\infty, f(\alpha_1), f(\alpha_2), f(\alpha_3), f(\alpha_4)) : f \in \mathbb{F}_m[X], \deg f \leq 1 \} \\ Y_2 &= \{ (f(\alpha_1), \infty, f(\alpha_2), f(\alpha_3), f(\alpha_4)) : f \in \mathbb{F}_m[X], \deg f \leq 1 \} \\ Y_3 &= \{ (f(\alpha_1), f(\alpha_2), \infty, f(\alpha_3), f(\alpha_4)) : f \in \mathbb{F}_m[X], \deg f \leq 1 \} \\ Y_4 &= \{ (f(\alpha_1), f(\alpha_2), f(\alpha_3), \infty, f(\alpha_4)) : f \in \mathbb{F}_m[X], \deg f \leq 1 \} \\ Y_5 &= \{ (f(\alpha_1), f(\alpha_2), f(\alpha_3), f(\alpha_4), \infty) : f \in \mathbb{F}_m[X], \deg f \leq 1 \} \end{split}$$

Clearly the sets Y_i are disjoint and have cardinality m^2 . Moreover, if elements $x, y \in Y_i$ agree on two components not including the *i*th, then x = y.

Define sets of words C_1, C_2, C_3, C_4 , and C_5 of length 5 over the alphabet ($\mathbb{F}_3 \times$ \mathbb{F}_m $\cup \{(\infty,\infty)\}$ by

$$\begin{split} C_i &= \{ ((x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4),(x_5,y_5)) : \\ &\quad (x_1,x_2,x_3,x_4,x_5) \in X_i \text{ and } (y_1,y_2,y_3,y_4,y_5) \in Y_i \} \end{split}$$

for all $i \in \{1, 2, 3, 4, 5\}$. Note that $|C_i| = |X_i| \times |Y_i| = 3m^2$.

CONSTRUCTION 4. Let q be of the form 3m + 1, where m is a prime power and $m \geq 4$. Define sets C_1, C_2, C_3, C_4 , and C_5 of words of length 5 over the alphabet $F = (\mathbb{F}_3 \times \mathbb{F}_m) \cup \{(\infty, \infty)\}$ as above. Then the code C defined by

$$C = C_1 \cup C_2 \cup C_3 \cup C_4 \cup C_5$$

is a 3-frameproof code of length 5 and cardinality $\frac{5}{3}q^2 - \frac{10}{3}q + \frac{5}{3}$. Proof. The subsets C_i are pairwise disjoint and $|C_i| = 3m^2 = \frac{1}{3}(q^2 - 2q + 1)$. Hence the code C has the claimed cardinality. It remains to show that C is 3-frameproof.

For a codeword $x \in C$, let $\pi_1(c)$ be the word in $X_1 \cup X_2 \cup X_3 \cup X_4 \cup X_5$ obtained by replacing each component $(a, b) \in F$ of c by the element $a \in \mathbb{F}_3 \cup \{\infty\}$. Note that $\pi_1(c) \in X_i$ if and only if $c \in C_i$. Similarly, define $\pi_2(c)$ to be the word in $Y_1 \cup Y_2 \cup Y_3 \cup Y_4 \cup Y_5$ obtained by replacing each component $(a, b) \in F$ of c by the element $b \in \mathbb{F}_m \cup \{\infty\}$.

Suppose $x \in C$ and let $P \subseteq C$ be such that $|P| \leq 3$ and $x \in \operatorname{desc}(P)$. We must show that $x \in P$. Now $x \in C_j$ for some $j \in \{1, 2, 3, 4, 5\}$. So the *j*th component of x is (∞, ∞) and $\pi_1(x) \in X_j$. Since $|P| \leq 3$, there exists $y \in P$ that agrees with x in 2 or more components other than the *j*th. We aim to show that x = y.

Since x and y agree in two or more components, the same is true for $\pi_1(x)$ and $\pi_1(y)$. Hence $\pi_1(y) = \pi_1(x)$. In particular, we have that $\pi_1(y) \in X_j$ and so $y \in C_j$.

Since $x, y \in C_j$, we have that $\pi_2(x), \pi_2(y) \in Y_j$. Moreover, since x and y agree in two components not including the *j*th, the same is true for $\pi_2(x)$ and $\pi_2(y)$. This implies that $\pi_2(x) = \pi_2(y)$. Since $\pi_1(x) = \pi_1(y)$ and $\pi_2(x) = \pi_2(y)$ we find that $x = y \in P$ as required. П

We remark that the condition $m \ge 4$ in the statement of Construction 4 can be weakened to $m \geq 3$ if we set $\alpha_4 = \infty$ in the definition of the sets Y_i . (See the remark after the statement of Construction 2.)

SIMON R. BLACKBURN

8. Discussion. Two questions suggest themselves for further work. First, can the upper bound of Corollary 9 be made more explicit by determining the constant $m(t, k, \ell)$ exactly in all cases? Erdős [7] warns that this does not seem easy. Second, is it the case that the upper bound of Corollary 9 is tight? The most tempting cases to consider are when we know the explicit value of $m(t, k, \ell)$ used in Corollary 2, namely when t = 1, t = 2, and $\ell = tk$. The case t = 1 occurs when $\ell = 1 \mod c$ and has already been dealt with by Corollary 5. The case t = 2 occurs when $\ell = 2 \mod c$. So is there a *c*-frameproof code of cardinality approximately $(\ell/(\ell - \lceil \ell/c \rceil))q^{\lceil \ell/c \rceil})$ when $\ell = 2 \mod c$? The case $\ell = tk$ occurs when ℓ is a multiple of *c*. So is there a *c*-frameproof code of cardinality approximately $cq^{\ell/c}$ when ℓ is a multiple of *c*?

Acknowledgments. Many thanks to Jessica Staddon and Rob Waiser for their careful readings of an earlier manuscript. Thanks to Dominic Welsh and Lucia Moura for pointing me in the direction of some literature on intersecting set systems, and thanks to Peter Wild for some useful discussions. Also, thanks to an anonymous referee for suggested improvements to the paper (including a strengthening of Corollaries 4 and 5) and for pointing out the reference [10].

REFERENCES

- [1] I. ANDERSON, Combinatorics of Finite Sets, Oxford University Press, Oxford, UK, 1987.
- B. BOLLOBÁS, D.E. DAYKIN, AND P. ERDŐS, Sets of independent edges in a hypergraph, Quart. J. Math. Oxford Ser. 2, 27 (1976), pp. 25–32.
- D. BONEH AND J. SHAW, Collision-secure fingerprinting for digital data, IEEE Trans. Inform. Theory, 44 (1998), pp. 1897–1905.
- [4] B. CHOR, A. FIAT, AND M. NAOR, *Tracing traitors*, in Advances in Cryptology—CRYPTO '94, Y.G. Desmedt, ed., Lecture Notes in Comput. Sci. 839, Springer, Berlin, 1994, pp. 257–270.
- [5] G. COHEN AND S. ENCHEVA, Efficient constructions of frameproof codes, Electron. Lett., 36 (2000), pp. 1840–1842.
- M. DEZA AND P. FRANKL, Erdős-Ko-Rado theorem—22 years later, SIAM J. Algebraic Discrete Methods, 4 (1983), pp. 419–431.
- [7] P. ERDŐS, A problem on independent r-tuples, Ann. Univ. Sci. Budapest Eőtvős Sect. Math., 8 (1965), pp. 93–95.
- [8] P. ERDŐS, C. KO, AND R. RADO, Intersection theorems for systems of finite sets, Quart. J. Math. Oxford Ser. 2, 12 (1961), pp. 313–320.
- [9] A. FIAT AND T. TASSA, Dynamic traitor tracing, in Advances in Cryptology—CRYPTO '99, M. Weiner, ed., Lecture Notes in Comput. Sci. 1666, Springer, Berlin, 1999, pp. 354–371.
- [10] H.-D.O.F. GRONAU, An extremal problem for set families, Ann. Inst. Mat. Univ. Nac. Autonoma Mexico, 25 (1985), pp. 1–10.
- [11] G.O.H. KATONA, A simple proof of the Erdős-Ko-Rado theorem, J. Combin. Theory Ser. B, 13 (1972), pp. 183–184.
- [12] J.N. STADDON, D.R. STINSON, AND R. WEI, Combinatorial properties of frameproof and traceability codes, IEEE Trans. Inform. Theory, 47 (2001), pp. 1042–1049.
- [13] D.R. STINSON AND R. WEI, Combinatorial properties and constructions of traceability schemes and frameproof codes, SIAM J. Discrete Math., 11 (1998), pp. 41–53.