A tight bound for frameproof codes viewed in terms of separating hash families

Tran van Trung
Institut für Experimentelle Mathematik
Universität Duisburg-Essen
Ellernstrasse 29
45326 Essen, Germany
`trung@iem.uni-due.de`

Abstract

Frameproof codes have been introduced for use in digital fingerprinting that prevent a coalition of w or fewer legitimate users from constructing a fingerprint of another user not in the coalition. It turns out that w-frameproof codes are equivalent to separating hash families of type $\{1, w\}$. In this paper we prove a tight bound for frameproof codes in terms of separating hash families.

1 Introduction

Let Q be a finite set of size q and let N be a positive integer. A subset $C \subseteq Q^N$ with $|C| = n$ is called C an (N, n, q) code. The elements of C are called codewords. Each codeword $x \in C$ is of the form $x = (x_1, \ldots, x_N)$, where $x_i \in Q$, $1 \leq i \leq N$. For any subset of codewords $P \subseteq C$, the set of *descendants* of P, denoted $\mathrm{desc}(P)$, is defined by

$$\mathrm{desc}(P) = \{x \in Q^N : x_i \in \{a_i : a \in P\},\ 1 \leq i \leq N\}.$$

Let C be an (N, n, q) code and let $w \geq 2$ be an integer. C is called a *w-frameproof code* if for all $P \subseteq C$ with $|P| \leq w$, we have that $\mathrm{desc}(P) \cap C = P$. Frameproof codes were first introduced by Boneh and Shaw [6], for use in fingerprinting of digital data to prevent a coalition of w or fewer legitimate users from constructing a copy of fingerprint of another user not in the coalition. Frameproof codes and their applications have been studied extensively, see for instance, [6], [13], [9], [16], [18], [15], [4], [10]. One of the basic problems is the studying of upper bounds on the cardinality of frameproof codes. Many strong bounds have been obtained in the papers [16], [15], [4]. It turns out that frameproof codes are a special type of separating hash families (SHF). Let h be a function from a set X to a set Y and let $C_1, C_2, \ldots, C_t \subseteq X$ be t pairwise disjoint subsets. We say that h *separates* $C_1, C_2, \ldots, C_t$ if $h(C_1), h(C_2), \ldots, h(C_t)$ are pairwise disjoint, where $h(C_i) = \{h(x) \mid x \in C_i\}$. Let $|X| = n$ and $|Y| = q$. We call a set $\mathcal{H}$ of N functions from X to Y an $(N; n, q, \{w_1, \ldots, w_t\})$-*separating hash family*, denoted by $\mathsf{SHF}(N; n, q, \{w_1, \ldots, w_t\})$, if for all pairwise disjoint subsets $C_1, \ldots, C_t \subseteq X$ with $|C_i| = w_i$, for $i = 1, \ldots, t$, there exists at least one function $h \in \mathcal{H}$ that separates $C_1, C_2, \ldots, C_t$. The multiset $\{w_1, w_2, \ldots, w_t\}$ is the *type* of the separating hash family. Separating hash families provide a link to many known combinatorial structures such as perfect hash families, frameproof codes, secure frameproof codes, identifiable parent property codes. Many results on separating hash families can be found in [18], [19], [5], [20], [14], [1], [2], [3], [11].

Frameproof codes and separating hash families have the following connection. An (N, n, q) w-frameproof codes exists if and only if an $\mathsf{SHF}(N; n, q, \{1, w\})$ exists. As it is more convenient to work with separating hash families, we will prove the results in this paper in terms of separating hash families.

It is often useful to present an $\mathsf{SHF}(N; n, q, \{w_1, \ldots, w_t\})$ as an $N \times n$ matrix on q symbols, say A. The rows of A correspond to the hash functions in the family, the columns correspond to the elements in the domain X, and the entry in row f and column x is $f(x)$. We call A the matrix representation of the hash family. The matrix A has the following property. For given disjoint sets of columns $C_1, C_2, \ldots, C_t$ with $|C_i| = w_i$, $1 \leq i \leq t$, there exists at least one row f of A such that

$$\{\mathsf{A}(f, x) : \ x \in C_i\} \cap \{\mathsf{A}(f, x) : \ x \in C_j\} = \emptyset,$$

for all $i \neq j$, i.e. row f separates the column sets $C_1, C_2, \ldots, C_t$. Now if we write the codewords of an (N, n, q) w-frameproof code columnwise as an $N \times n$ matrix A, i.e. each codeword is a column of A, then A is the matrix representation of an $\mathsf{SHF}(N; n, q, \{1, w\})$. The problem of determining an upper bound on the cardinality of an (N, n, q) w-frameproof code becomes the problem of determining an upper bound on the number of columns of A for given N, q, and w. When $N \leq w$, it has been shown that $n \leq w(q-1)$, see [16], or [4]. The more interesting case is when $N > w$. Strong bounds for case $N > w$ are obtained in [16], [15], [4], [3].

We are interested in the case $N = wd + 1$ with $d \geq 1$. Several previously strong bounds known for this case are found in [4], [3].

Theorem 1 ([4]) *Let N, q, w and d be positive integers such that $N = wd + 1$, $w \geq 2$. Suppose there is an (N, n, q) w-frameproof code. Then $n \leq q^{d+1} + O(q^d)$.*

Theorem 2 ([3]) *Let N, q and d be positive integers such that $N = 2d + 1$. Suppose there is an (N, n, q) 2-frameproof code. Then $n \leq q^{d+1}$.*

Theorem 3 ([3]) *Let q and w be positive integers such that $w \geq 2$, $q \geq w + 1$. Suppose there is an $(w + 1, n, q)$ w-frameproof code. Then $n \leq q^2$.*

The bounds in Theorems 2, 3 are tight.

The aim of the paper is to prove the following bound on the cardinality of w-frameproof codes.

Theorem 4 *Let d, q, w be positive integers such that $q \geq w \geq 2$. Suppose there exists an (N, n, q) w-frameproof code with $N = wd + 1$. Then $n \leq q^{d+1}$.*

The bound of Theorem 4 is tight as shown in the next section.

2 A tight bound for separating hash families of type $\{1, w\}$

For the sake of completeness we include the following simple lemma.

Lemma 1 *An (N, n, q) w-frameproof code is equivalent to an $\mathsf{SHF}(N; n, q, \{1, w\})$.*

Proof. Let A be an $N \times n$ matrix having entries from a set of q symbols. Let $\{c\}$ and P be any given disjoint subsets of columns of A with $|\{c\}| = 1$ and $|P| \leq w$, where w is an integer such that $w \geq 2$. We may view A as an (N, n, q) code whose codewords are the columns. Assume that A is an (N, n, q) w-frameproof code. By definition this is equivalent to $\mathrm{desc}(P) \cap \mathsf{A} = P$. Further, $\mathrm{desc}(P) \cap \mathsf{A} = P$ is equivalent to the fact that there is a row i that separates $\{c\}$ and P. The latter says that A is the matrix representation of an $\mathsf{SHF}(N; n, q, \{1, w\})$. $\square$

Let A be the matrix representation of an $\mathsf{SHF}(wd + 1; n, q, \{1, w\})$, where d is a positive integer. Thus A is an $N \times n$ matrix with $N = wd + 1$ having entries from a set of q symbols. Let $1, 2, \ldots, N$ denote the row positions of A. Note that the rows of A may be permuted but the row positions are fixed. Consider two partitions of the row positions of A.

The first partition, denoted by $R_1, R_2, \ldots, R_w$, is defined by

$$R_1 = \{1, \ldots, d, d + 1\},\ R_2 = \{d + 2, \ldots, 2d + 1\},\ \ldots\ ,\ R_w = \{(w - 1)d + 2, \ldots, wd + 1\}.$$

So we have $|R_1| = d + 1$ and $|R_2| = \cdots = |R_w| = d$.

The second partition, denoted by $Z_1, Z_2, \ldots, Z_w$, is defined by

$$Z_1 = \{1, \ldots, d\}\ ,\ Z_2 = \{d + 1, \ldots, 2d\}\ ,\ \ldots\ ,\ Z_{w-1} = \{(w - 2)d + 1, \ldots, (w - 1)d\}\ ,$$
$$Z_w = \{(w - 1)d + 1, \ldots, wd + 1\}.$$

So we have $|Z_1| = \cdots = |Z_{w-1}| = d$ and $|Z_w| = d + 1$.

Let c_i be a column of A. We write $c_i = c_{i1}||c_{i2}||\ldots||c_{iw}$ (resp. $c_i = c'_{i1}||c'_{i2}||\ldots||c'_{iw}$) where c_{ij} (resp. c'_{ij}) is the restriction of c_i to the row positions of R_j (resp. of Z_j). Thus c_{ij} and c'_{ij} are a d-tuple or a $d + 1$-tuple of symbols.

Using the notation just described we first prove the following lemma.

Lemma 2 *Let A be the matrix representation of an $\mathsf{SHF}(wd + 1; n, q, \{1, w\})$, where d is a positive integer. Suppose that there are two columns c_1 and c_2 of A agreeing in the first $(d + 1)$ row positions of R_1 (resp. in the last $(d + 1)$ row positions of Z_w). Then each of the columns c_1 and c_2 has at least a unique d-tuple corresponding to one of $R_2, \ldots, R_w$ (resp. to one of $Z_1, \ldots, Z_{w-1}$).*

Proof. By using the notation described above we write $c_i = c_{i1}||c_{i2}||\ldots||c_{iw}$ for $i = 1, 2$, where c_{ij} is the restriction of c_i to R_j. Since c_1 and c_2 agree in R_1, we have that $c_{11} = c_{21}$, where c_{11} and c_{21} are $d + 1$-tuples of symbols. Whereas $c_{i2}, \ldots, c_{iw}$ are all d-tuples of symbols. Assume, by contradiction, that all d-tuples $c_{12}, \ldots, c_{1w}$ are repeated in $R_2, \ldots, R_w$, say in columns $s_2, \ldots, s_w$. Then we have the following configuration in A.

$$\begin{array}{c|ccccccc|c}
\cline{2-8}
 & \dots & c_1 & c_2 & s_2 & \dots & s_w & \dots & \\
\cline{2-8}
R_1 \to & \dots & c_{11} & c_{21} & * & \dots & * & \dots & \\
R_2 \to & \dots & c_{12} & c_{22} & c_{12} & \dots & * & \dots & \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \mathsf{A} \\
R_w \to & \dots & c_{1w} & c_{2w} & * & \dots & c_{1w} & \dots & \\
\cline{2-8}
\end{array}$$

But then the two column sets $\{c_1\}$ and $\{c_2, s_2, \dots, s_w\}$ cannot be separated in A, a contradiction. Hence, at least one of the d-tuples $c_{12}, \dots, c_{1w}$ must be unique. A similar argument shows that at least one of the d-tuples $c_{22}, \dots, c_{2w}$ is unique. When columns c_1 and c_2 agree in the last $(d+1)$ row positions, we obtain the statement with a similar argument by using partition $Z_1, \dots, Z_w$. $\square$

We now prove Theorem 4 in terms of separating hash families, which is equivalent to the following theorem.

Theorem 5 *Let q, w, and d be positive integers such that $q \geq w \geq 2$. Suppose that there exists an $\mathsf{SHF}(wd+1; n, q, \{1, w\})$. Then $n \leq q^{d+1}$.*

Proof. Let A be the $(wd+1) \times n$ - matrix representation of an $\mathsf{SHF}(wd+1; n, q, \{1, w\})$. The idea of the proof is to show that if $n \geq q^{d+1}+1$, then there are q^d unique d-tuples of symbols corresponding to R_w or Z_1, by using Lemma 2 and by permuting the rows of A. This leads to a contradiction, as there are no free d-tuples available to fill the columns of A.

Now assume, by contradiction, that $n = q^{d+1}+1$. We focus on R_1, R_w and Z_1, Z_w. The proof consists of a finite number of repeated steps, which prove that the number of unique d-tuples of symbols corresponding to R_w and Z_1 strictly increases with the number of steps. More precisely, each step begins with u_1 pairs of columns agreeing in the $(d+1)$ rows of R_1 and ends in x unique d-tuples corresponding to R_w, y unique d-tuples corresponding to Z_1, and u_2 pairs of columns agreeing in the $(d+1)$ rows of R_1 with $u_2 > u_1$. During each step the rows corresponding to R_1, R_w, Z_1 and Z_w have usually been changed. To illustrate the idea we show the first two steps.

Step 1.

Since $n = q^{d+1}+1$, there are two columns c_1 and c_2 of A agreeing in the $(d+1)$ rows of R_1. By permuting the d-tuples $c_{12}, \dots, c_{1w}$ if necessary, we may assume by using Lemma 2 that c_{1w} is a unique d-tuple. This is because, if c_{1j} is a unique d-tuple with $j \neq w$, we interchange the rows in R_j and in R_w in such a way that c_{1j} becomes c_{1w} of R_w. Since this type of permuting rows in A will be repeated frequently, we say for short that we *update* the rows of R_w (by using the rows of R_j). Note that permuting the rows of A does not effect the separation property of A. Since c_{1w} is unique, the maximal number of remaining d-tuples corresponding to R_w is $q^d - 1$. If each of these d-tuples appears at most q times in R_w, we can fill only $(q^d-1)q$ columns of A. So there are $q^{d+1} - (q^d-1)q = q$ columns, whose d-tuples are repeated at least $q+1$ times in R_w. Thus there are at least q $(d+1)$-tuples of symbols repeated in Z_w, because there are q symbols altogether. Each of these q repeated $(d+1)$-tuples gives at least one unique d-tuple distributed in the rows of $Z_1, \dots, Z_{w-1}$. Since $q \geq w > w-1$, at least one Z_i, $i \in \{1, \dots, w-1\}$ contains at least 2 unique d-tuples. If $i \neq 1$, then by updating the rows of Z_1 we may assume that Z_1 contains 2 unique d-tuples. The (maximal) remaining $q^d - 2$ d-tuples in Z_1 are distributed in th $q^{d+1} + 1 - 2 = q^{d+1} - 1$ columns of A. Again if each of these $q^d - 2$ d-tuples appears at most q times, we can fill at most $(q^d - 2)q$ columns. So there are $q^{d+1} + 1 - 2 - (q^d - 2)q = 2q - 1$ columns with d-tuples in Z_1 that have to repeat at least $q + 1$ times. Thus there are at least $2q - 1$ repeated $(d+1)$-tuples in the $(d+1)$ rows of R_1.

Step 2.

From Step 1 we have that there are at least $2q - 1$ pairs of columns such that each pair agrees in the $(d+1)$ rows of R_1. By using Lemma 2 each of these pairs provides at least one unique d-tuple distributed in $R_2, \ldots, R_w$. Since $2q - 1 \geq 2w - 1 > 2(w-1)$, there is an R_i that contains at least 3 unique d-tuples. If $R_i \neq R_w$, we update the rows of R_w. So we may assume that R_w contains at least 3 unique d-tuples. Hence there are at most $q^d - 3$ remaining d-tuples in R_w distributed in $(q^{d+1} + 1 - 3) = (q^{d+1} - 2)$ columns of A. If each of these d-tuples appears at most q times in R_w, then only $(q^d - 3)q$ columns of A can be filled. So there are $(q^{d+1} + 1) - 3 - (q^d - 3)q = 3q - 2$ columns, whose d-tuples are repeated at least $q + 1$ times in R_w. Hence there are at least $3q - 2$ $(d+1)$-tuples repeated in Z_w. Since each pair of these repeated columns provides at least one unique d-tuple in some Z_i, $i \in \{1, \ldots, w-1\}$, we have at least $3q - 2$ unique d-tuples distributed in $Z_1, \ldots, Z_{w-1}$. Since $3q - 2 \geq 3w - 2 > 3(w-1)$, there is an Z_i that contains at least 4 unique d-tuples. If $i \neq 1$, then again by updating the rows of Z_1 we may assume that Z_1 contains 4 unique d-tuples. The (maximal) remaining $q^d - 4$ d-tuples in Z_1 are distributed in the remaining $q^{d+1} + 1 - 4 = q^{d+1} - 3$ columns of A. Again if each of these d-tuples appears at most q times, we can fill at most $(q^d - 4)q$ columns. So there are $q^{d+1} + 1 - 4 - (q^d - 4)q = 4q - 3$ columns with d-tuples in Z_1 that have to repeat at least $q + 1$ times. Hence there are at least $4q - 3$ repeated $(d+1)$-tuples in the $(d+1)$ rows of R_1.

When repeating the argument as shown in the two steps above, we see that the number of unique d-tuples in R_w and Z_1 strictly increases with the number of steps. More precisely, at Step i with $i \leq q^d/2$ we have that R_w contains $(2i - 1)$ unique d-tuples and Z_1 contains $2i$ unique d-tuples. So, if q is even, Z_1 (with its rows updated) contains all q^d unique d-tuples at step $i = q^d/2$. If q is odd, R_w (with its rows updated) contains all q^d unique d-tuples at step $i = \lceil q^d/2 \rceil + 1$. This shows that there are no more free d-tuples in R_w or in Z_1 to fill the columns of A after a finite number of steps. This contradiction completes the proof. $\square$

To show that the bound of Theorem 5 is tight we make use of a combinatorial structure called orthogonal arrays. An *orthogonal array* $\mathsf{OA}(t, N, m)$ is an $N \times m^t$ array A with entries from a set of $m \geq 2$ symbols such that within any t rows of A every possible $t-$tuple of symbols occurs exactly once. This property is equivalent to the fact that every two columns of A agree in at most $t - 1$ rows, see for example [12]. The just given definition of orthogonal arrays is in fact the definition of orthogonal arrays of *strength* t and *index* 1. A classical construction of orthogonal arrays is as follows [8]. Let q be a prime power and $t \geq 2$. Let $\mathcal{P} = \{P_1, P_2, \ldots, P_{q^t}\}$ be the set of all polynomials of degree at most $t - 1$ over the finite field $\mathbb{F}_q$. Now let $\mathcal{R}$ be a subset of elements of $\mathbb{F}_q \cup \{\infty\}$. Define an $|\mathcal{R}| \times q^t$ array A in which the entry $\mathsf{A}(u, j)$ is $P_j(u)$ if $u \in \mathcal{R} \setminus \{\infty\}$ and is a_{t-1} when $P_j(x) = \sum_{i=0}^{t-1} a_i x^i$ and $u = \infty$. Then A is an $\mathsf{OA}(t, |\mathcal{R}|, q)$. For more about orthogonal arrays we refer the reader to [12].

As an application of orthogonal arrays we obtain the following theorems showing that the bound of Theorem 5 is tight.

Theorem 6 *Let q, d, w be positive integers such that q is a prime power with $q \geq wd$ and $w \geq 2$. Then there exists an* $\mathsf{SHF}(wd+1; q^{d+1}, q, \{1, w\})$.

Proof. Let q be a prime power such that $q \geq wd$. Let $\mathcal{R} \subseteq \mathbb{F}_q \cup \{\infty\}$ with $|\mathcal{R}| = wd+1$. Consider the classical orthogonal array $\mathsf{OA}(d+1, |\mathcal{R}|, q)$ which is an $(wd+1) \times q^{d+1}$ array A. Now any two different columns of A agree in at most d rows. It follows that for given two disjoint subsets of columns C_1 and C_2 of A with $|C_1| = 1$ and $|C_2| = w$, there is at least one row that separates C_1 and C_2. Hence A is an $\mathsf{SHF}(wd+1; q^{d+1}, q, \{1, w\})$. $\square$

When q is not a prime power, we have the following result.

Theorem 7 *Let $q = p_1^{e_1} p_2^{e_2} \ldots p_s^{e_s}$ be a prime power factorization of an integer q with $q \geq 2$ such that $p_1^{e_1} < p_2^{e_2} < \ldots < p_s^{e_s}$. Let w and d be positive integers such that $p_1^{e_1} \geq wd$ and $w \geq 2$. Then there exists an* $\mathsf{SHF}(wd+1; q^{d+1}, q, \{1, w\})$.

Proof. It is known by a result of Bush (see [7] or [12], 7.20 Theorem, page 226) that there is an $\mathsf{OA}(d+1, k, q)$ for $d+1 < p_1^{e_1}$ and $k \leq p_1^{e_1} + 1$. If we choose $k = wd+1$, then an $\mathsf{OA}(d+1, wd+1, q)$ provides an $\mathsf{SHF}(wd+1; q^{d+1}, q, \{1, w\})$. $\square$

References

[1] M. Bazrafshan and Tran van Trung, Bounds for separating hash families, *J. Combin. Theory Ser. A* 118 (2011), 1129–1135.

[2] M. Bazrafshan, Separating Hash Families, PhD thesis, University of Duisburg-Essen, 2011.

[3] M. Bazrafshan and Tran van Trung, Improved bounds for separating hash families, *Des. Codes Cryptpgr.* DOI 10.1007/s10623-012-9673-7 (2012).

[4] S. R. Blackburn, Frameproof codes, *SIAM J. Discrete Math.*, Vol.16, No. 3 (2003), 499–510.

[5] S. R. Blackburn, T. Etzion, D. R. Stinson and G. M. Zaverucha, A bound on the size of separating hash families, *J. Combin. Theory Ser. A* 115 (2008), 1246–1256.

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

[7] K. A. Bush, A generalization of a theorem due to MacNeish, *Ann. Math. Stat.* 23 (1952) 293–295.

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

[9] , B. Chor, A. Fiat and M. Naor, Tracing traitors, in Advances in Cryptology - CRYPTO'94, Y. G. Desmedt, ed., *Lecture Notes in Computer Science*, 839, Springer, Berlin (1994), 257 – 270

[10] C. J. Colbourn, D. Horsley, and V. R. Syrotiuk, Frameproof codes and compressive sensing, *Forty-Eighth Annual Allerton Conference*, Allerton House, UIUC, Illinois, USA, September 29 - October 1, 2010, 985–990.

[11] C. J. Colbourn, D. Horsley, and C. McLean, Compressive sensing matrices and hash families, *Transactions on Communications* Vol. 59, Nr.7, July 2011, 1840–1845.

[12] C. J. Colbourn and J. H. Dinitz, editors. *The CRC Handbook of Combinatorial Designs* Chapman and Hall/CRC, Boca Raton, FL, 2nd edition, 2007.

[13] 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), 354–371.

[14] P. C. Li, R. Wei and G. H. J. van Rees, Constructions of 2-cover-free families and related separating hash families, *J. Combin. Des.* 14 (2006), 423–440.

[15] P. Sarkar, D. R. Stinson, Frameproof and IPP codes, Progress in Cryptology – Indocrypt 2001, *Lecture Notes in Computer Science*, Springer, Vol.2247, (2001), 117–126.

[16] J. N. Staddon, D. R. Stinson and R. Wei, Combinatorial properties of frameproof and traceability codes, *IEEE Transaction on Information Theory* 47 (2001), 1042-1049.

[17] D. R. Stinson and R. Wei, Combinatorial properties and constructions of traceability schemes and frameproof codes, *SIAM J. Discrete Math.* 11 (1998), 41-53.

[18] D. R. Stinson, Tran van Trung and R. Wei, Secure frameproof codes, key distribution patterns, group testing algorithms and related structures, *J. Statist. Plann. Inference* 86 (2000), 595–617.

[19] D. R. Stinson, R. Wei and K. Chen, On Generalized Separating Hash Families, *J. Combin. Theory Ser. A* 115 (2008), 105-120.

[20] D. R. Stinson, G. M. Zaverucha, Some improved bounds for secure frameproof codes and related separating hash families, *IEEE Transaction on Information Theory* 54 (2008), 2508–2514.