New Bounds for Generalized Separating Hash Families

Douglas R. Stinson* and Gregory M. Zaverucha
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo ON, N2L 3G1, Canada
{dstinson,gzaveruc}@uwaterloo.ca

June 15, 2007

Abstract

The main result of this paper is a necessary condition for generalized separating hash families. We extend previous methods used to obtain upper bounds for separating hash family types $\{w, w\}$ and $\{w, w-1\}$ to the general case $\{w_1, w_2, \ldots, w_t\}$ for $t \geq 2$.

1 Introduction

Let $\mathcal{F}$ be a set of N functions $f_i : X \to Y$ where X, Y are sets with $|X| = n$ and $|Y| = m$. We call $\mathcal{F}$ an $(N; n, m)$-hash family. For disjoint sets $C_1, \ldots, C_t \subset X$ with cardinalities $w_1, \ldots, w_t$, define the multisets $B_{i,j} = \{f_i(c) : c \in C_j\}$ (the result of "hashing" the set C_j with the function f_i). If there exists some f_j such that $\bigcap B_{i,j} = \emptyset$, then $\mathcal{F}$ is an $(N; m, n)$-*separating hash family* of *type* $\{w_1, \ldots, w_t\}$. A more compact notation we will use is $\mathrm{SHF}(N; n, m, \{w_1, \ldots, w_t\})$. In the special case $w_i = 1$ for all i, $\mathcal{F}$ is called a *perfect hash family*.

The matrix representation of a hash family will prove to be the most useful representation for this paper. Given an $(N; n, m)$-hash family, first index X by the natural numbers $\{1, \ldots, n\}$. Then an $N \times n$ matrix A is naturally defined as $A_{i,j} = f_i(X_j)$. To get an upper bound on n (in terms of N and m), the general strategy involves showing that a particular choice of n means that A always contains a submatrix which is impossible in an SHF. Such a submatrix is referred to as a *forbidden configuration*.

Many applications of separating hash families come when each column of A is a viewed as a codeword in an (n, N, m)-code. A survey of earlier (pre-1994) work in this area is Sagalovich [2], while Stinson, Wei and Chen [5] list many

*Research supported by NSERC grant 203114-06

newer papers. For cryptographic applications of separating hash families related to fingerprinting digital data, see Barg, Blakley and Kabatiansky [10].

We will first re-prove a bound for SHF of type $\{w, 1\}$ in Section 2, to illustrate part of the method used to prove the main result. The idea in this proof is then combined with the concept of *staircases* (defined in [11]), to obtain previously unknown bounds for the small types $\{5, 2\}, \{5, 3\}$. Section 4 generalizes this to SHF of type $\{w, d\}$. The result for type $\{w, d\}$ leads to a necessary condition for SHF of type $\{w_1, \ldots, w_t\}$, presented in Section 4.

1.1 Related Work

The following two theorems follow easily from the definition of separating hash families, but are quite useful.

Theorem 1.1. *Suppose A is an $SHF(N; n, m, \{w_1, w_2, \ldots, w_t\})$, and let $w_1' \leq w_1$. Then A is also an $SHF(N; n, m, \{w_1', w_2, \ldots, w_t\})$.*

Theorem 1.2. *Suppose A is an $SHF(N; n, m, \{w_1, w_2, \ldots, w_t\})$ and define $w_1' = w_1 + w_2$. Then A is an $SHF(N; n, m, \{w_1', w_3, \ldots, w_t\})$.*

The next lemma allows us to prove a bound for a fixed number of rows, then "group rows" to obtain a bound for arbitrary N.

Lemma 1.3. *Suppose there exists an $SHF(N; n, m, \{w_1, \ldots, w_t\})$ and let $c \geq 2$ be an integer. Then there exists an $SHF(\lceil \frac{N}{c} \rceil; n, m^c, \{w_1, \ldots, w_t\})$.*

SHF of small types were studied by Stinson, Wei and Chen [5]. Some of their results were later extended to general types $\{w, w\}$ and $\{w, w-1\}$ by Stinson and Zaverucha [11]. Types $\{w, w\}$ are also known as *secure frameproof codes*, and have been used in data fingerprinting applications. The best known upper bound on n for SHF$(N; n, m, \{w, w\})$ is the following theorem.

Theorem 1.4 (Stinson, Zaverucha [11])**.** *If an $SHF(N; n, m, \{w, w\})$ exists, it holds that*

$$n \leq m^{\lceil \frac{N}{2w-1} \rceil} + (w-1)(2w-1)(m^{\lceil \frac{N}{2w-1} \rceil} - 1) \, .$$

The proof of Theorem 1.4 works by showing that when n is large enough, the matrix representation of the hash family must contain a submatrix isomorphic to a staircase matrix. This pattern gives a contradiction, since the even and odd indexed columns of the staircase cannot be separated. We will use this forbidden configuration as part of the proof for SHF of type $\{w, d\}$.

Independent of this work, Blackburn [12] also discovered a bound for SHF of type $\{w_1, \ldots, w_t\}$. The theorem is stated below. We omit the proof, which is graph theoretic and inspired by one of his earlier papers [8].

Theorem 1.5 (Blackburn [12])**.** *Suppose an $SHF(N; n, m, \{w_1, \ldots, w_t\})$ exists. Define $u = \sum_{i=1}^{t} w_i$. Then*

$$n \leq \binom{u}{2} m^{\lceil N/(u-1) \rceil} \, . \tag{1}$$

This bound has the same exponent as the result in Theorem 5.1, which is the best possible (Blackburn justifies this claim in [12]). In Section 5 we compare the two bounds in greater detail and show that the bound of Theorem 5.1 is stronger for all choices of w_i since the coefficient on m is smaller.

2 SHF of type $\{w, 1\}$

In this section we prove a necessary condition for SHF of type $\{w, 1\}$, not because it is new (see [6] or [13]), but to illustrate part of the proof method used in Section 4 for SHF of type $\{w, d\}$.

Theorem 2.1. *In an $SHF(w; n, m, \{w, 1\})$, it holds that $n \leq m+(w-1)(m-1)$.*

Proof. Let $N = w$, and suppose A is an SHF$(N; n, m, \{w, 1\})$ with $n = m+1+(N-1)(m-1)$. We will create a series of submatrices $A_{N-1} \subset \ldots \subset A_1 \subset A$. Let $A_0 = A$. Submatrix A_i has the property that all elements in the i-th row repeat at least twice. When we create A_1 from A, we delete no more than $m-1$ columns, since if all m symbols appeared exactly once, A would have m columns, but we assumed it has $m+1+(N-1)(m-1)$. Thus

$$|A_1| \geq |A_0| - (m-1) = m + 1 + (N-2)(m-1).$$

Assume that

$$|A_i| \geq |A_{i-1}| - (m-1) = m + 1 + (N-(i+1))(m-1). \tag{2}$$

When we create A_{i+1}, we delete at most $m-1$ columns since we assumed A_i has more than m columns. Then

$$\begin{aligned} |A_{i+1}| &\geq |A_i| - (m-1) \\ &= m + 1 + (N-(i+1))(m-1) - (m-1) \\ &= m + 1 + (N-(i+2))(m-1) \end{aligned}$$

and we conclude that (2) holds. The last submatrix A_{N-1}, has $m+1$ columns so we can be sure that at least one element repeats twice in row N. We permute columns to place the repeated element in the last two columns, and label elements in the rightmost column.

$*$	a
$\vdots$	$\vdots$
$*$	x
$*$	y
z	z

Now we know that each element in the last column will be repeated in the same row as we move from submatrix A_i to submatrix A_{i-1}. When each repetition occurs in a different column, A has a submatrix isomorphic to

a	$*$	$*$	$*$	$*$	a
	$\ddots$				$\vdots$
$*$	$*$	x	$*$	$*$	x
$*$	$*$	$*$	y	$*$	y
$*$	$*$	$*$	$*$	z	z

,

in which the first w columns cannot be separated from the last. If some repetitions occur in the same column, say all repetitions appear in $t < w$ columns, then A is not an SHF of type $\{t, 1\}$, and can therefore not be an SHF of type $\{w, 1\}$ by the contrapositive of Theorem 1.1. We have shown that A is not an SHF of type $\{w, 1\}$, a contradiction. □

We now use Lemma 1.3 to extend the result to an arbitrary number of rows. We will use this technique repeatedly in this paper, and will subsequently omit the proofs since they are analogous to the one below.

Corollary 2.2. *In an* $SHF(N; n, m, \{w, 1\})$, $n \leq wm^{\lceil \frac{N}{w} \rceil} - w + 1$

Proof. Suppose there exists an SHF$(N; n, m, \{w, 1\})$ where $n = wm^{\lceil \frac{N}{w} \rceil} - w$. Then by Lemma 1.3, with $c = w$ there exists an SHF$(w; wm' - w, m', \{w, 1\})$ with $m' = m^{\lceil \frac{N}{w} \rceil}$ which contradicts the previous theorem. □

Note that this bound is weaker (by one) than the bound of Staddon, Stinson and Wei [13]. They show that $n \leq wm^{\lceil \frac{N}{w} \rceil} - w$ holds in an SHF$(N; n, m, \{w, 1\})$.

3 SHF of type $\{5, d\}$

In this section we examine necessary conditions for the existence of SHF of type $\{5, d\}$ for $1 \leq d \leq 5$. Previous results were known for types $\{5, 1\}, \{5, 4\}$, and $\{5, 5\}$, while the bounds for the case $d = 2, 3$ are new. Proofs of the new bounds will illustrate the method used to obtain an upper bound for type $\{w, d\}$ in Section 4.

Theorem 3.1. *If an* $SHF(7; n, m, \{5, 3\})$ *exists, then* $n \leq m + 24(m - 1)$.

Proof. Suppose A is an SHF$(7; n, m, \{5, 3\})$ with $n = m+1+24(m-1)$. We will construct submatrices $A_6 \subset A_5 \subset \ldots \subset A_1 \subset A$, by deleting certain columns.

We construct A_1 so that elements appearing in the first row appear at least 8 times in the first row. The t_i elements appearing exactly i times, for $1 \leq i \leq 7$, will be deleted. It must be that $t_1 + \ldots + t_7 \leq 7(m-1)$ since if $t_1 + \ldots + t_7 = 7m$ then A would have $7m$ columns, fewer than the number we assumed. After deleting columns from A, A_1 has at least $m + 1 + 17(m - 1)$ columns.

We now delete columns of A_1 to form A_2, where all elements in row 2 appear at least 7 times in row 2. We delete no more than $t_1+\ldots+t_6 \leq 6(m-1)$ columns so $|A_2| \geq m+1+11(m-1)$. Similarly A_3 must have elements repeating 6 or more times in row 3, and $|A_3| \geq m+1+6(m-1)$ since we delete at most $t_1+\ldots+t_5 \leq 5(m-1)$ columns. In A_4 elements must repeat at least 5 times, by similar reasoning $|A_4| \geq m+1+2(m-1)$.

In A_5 and A_6 elements appearing in rows 5 and 6 (resp.) must appear at least twice in these rows. Both times we delete no more than $(m-1)$ columns. Therefore, $|A_5| \geq m+1+(m-1)$ and $|A_6| \geq m+1$.

Now we will derive a contradiction, showing that A is not an SHF of type $\{5,3\}$. A_6 has at least $m+1$ columns so we can be sure that there is a repeated element in the 7th row. Call this element a, and permute columns of A_6 so that both a appear in the rightmost columns. Let b be the element in row 6 above the rightmost a. There is a second b in row 6, giving two possibilities

b	b
a	a

or

b	$*$	b
$*$	a	a

as a submatrix of A_6. Let c be the element in row 5 in the rightmost column. Since c appears at least twice in row 5, we have the following two possibilities,

c	c
b	b
a	a

or

$\star$	$\star$	$\star$	c
$*$	b	$*$	b
$*$	$*$	a	a

where one of $\star$ is the other c. In both cases, this submatrix is not an SHF of type $\{3,1\}$ since the rightmost column cannot be separated from the other three.

We suppose the widest case occurs and consider what happens in the next row as a submatrix of A_4. All elements in the 4th row appear at least 5 times (in the 4th row). Therefore A_4 has a submatrix isomorphic to

d	d	$*$	$*$	$*$
$*$	c	$*$	$*$	c
$*$	$*$	b	$*$	b
$*$	$*$	$*$	a	a

The elements in the 3rd row of A_3 appear at least 6 times, giving

e	e	$*$	$*$	$*$	$*$
$*$	d	d	$*$	$*$	$*$
$*$	$*$	c	$*$	$*$	c
$*$	$*$	$*$	b	$*$	b
$*$	$*$	$*$	$*$	a	a

A similar situation exists for A_2 and A_1, and we finally get

$$\begin{array}{|c|c|c|c|c|c|c|c|}
\hline
g & g & * & * & * & * & * & * \\
\hline
* & f & f & * & * & * & * & * \\
\hline
* & * & e & e & * & * & * & * \\
\hline
* & * & * & d & d & * & * & * \\
\hline
* & * & * & * & c & * & * & c \\
\hline
* & * & * & * & * & b & * & b \\
\hline
* & * & * & * & * & * & a & a \\
\hline
\end{array} \tag{3}$$

as a submatrix of A. Number the columns from right to left. The sets of columns $\{2, 3, 4, 6, 8\}$ and $\{2, 3, 4\}$ cannot be separated, therefore A is not an SHF of type $\{5, 3\}$.

We assumed the widest case would occur for the bottom 3 rows in columns 2,3,4. If the repeated elements occurred in fewer than 3 columns, then A would not even be an SHF of type $\{4, 3\}$ or $\{3, 3\}$, and the proof still holds. □

Again, from the theorem above and Lemma 1.3 we get a bound for SHF of type $\{5, 3\}$ with N rows.

Corollary 3.2. *If an* $SHF(N; n, m, \{5, 3\})$ *exists,* $n \leq 25m^{\lceil \frac{N}{7} \rceil} - 24$.

Note that Theorem 3.1 started by showing that A could not have a submatrix isomorphic to an SHF of type $\{3, 1\}$ (in the bottom 3 rows). Then a forbidden configuration for SHF of type $\{3, 2\}$ was added (in the top 3 rows). We can prove a similar result for type $\{5, 2\}$ by combining a forbidden configuration for type $\{4, 1\}$ on the bottom, with one for type $\{2, 1\}$ on the top.

Theorem 3.3. *If an* $SHF(6; n, m, \{5, 2\})$ *exists, then* $n \leq m + 14(m - 1)$.

Extending to N rows gives the following corollary.

Corollary 3.4. *If an* $SHF(N; n, m, \{5, 2\})$ *exists, then* $n \leq 15m^{\lceil \frac{N}{6} \rceil} - 14$.

Results for type $\{5, d\}$ $1 \leq d \leq 5$ are gathered up in Table 1.

Type	Bound	Source
$\{5, 1\}$	$n \leq 5m^{\lceil \frac{N}{5} \rceil} - 5$	[6] or [13]
$\{5, 2\}$	$n \leq 15m^{\lceil \frac{N}{6} \rceil} - 14$	Corollary 3.4
$\{5, 3\}$	$n \leq 25m^{\lceil \frac{N}{7} \rceil} - 24$	Corollary 3.2
$\{5, 4\}$	$n \leq 15m^{\lceil \frac{N}{8} \rceil} - 14$	[11]
$\{5, 5\}$	$n \leq 37m^{\lceil \frac{N}{9} \rceil} - 36$	[11]

Table 1: Summary of results for SHF of type $\{5, d\}$ where $1 \leq d \leq 5$.

4 Necessary condition for type $\{w, d\}$

In this section we generalize the method used in Section 3 for SHF of type $\{5, 2\}$ and $\{5, 3\}$. Our main result will be a necessary condition for the existence of $\mathrm{SHF}(N; n, m, \{w, d\})$.

Proving Theorem 4.1 is a matter of adapting the proof of Theorem 3.1 to have a variable number of rows. Recall the forbidden configuration in the proof for type $\{5, 3\}$ came from combining forbidden configurations for types $\{3, 1\}$ and $\{3, 2\}$, with one column overlapping. This generalizes as follows: to show the existence of a forbidden configuration for type $\{w, d\}$ (where $d \leq w$), combine one for type $\{d, d-1\}$ with one for type $\{w-d+1, 1\}$.

Theorem 4.1. *If an $SHF(w+d-1; n, m, \{w, d\})$ exists, then*

$$n \leq m + (2dw - w - 1)(m - 1).$$

Proof. Let $N_0 = 2d-2$, $N_1 = w-d+1$ and $N = N_0 + N_1 = w+d-1$. Suppose A is an $\mathrm{SHF}(N; n, m, \{w, d\})$ where $n = m + 1 + (2dw - w - 1)(m - 1)$. Then

$$|A| = m + 1 + \left(N_1 - 1 + N_0 N - \frac{N_0(N_0 - 1)}{2} \right) (m - 1)$$

Let $K = N_1 - 1 + N_0 N - N_0(N_0 - 1)/2$, then $|A| = m + 1 + (K)(m - 1)$.

We will create a series of submatrices of A, each of which satisfy one of two properties, as indicated.

$$\underbrace{A_{N-1} \subset \ldots \subset A_{N_0+1}}_{\text{Property (ii)}} \subset \underbrace{A_{N_0} \subset A_{N_0-1} \subset \ldots \subset A_1}_{\text{Property (i)}} \subset A_0 = A$$

Property (i): elements in row i of A_i, $1 \leq i \leq N_0$, repeat at least $N - (i-2)$ times, and we claim that

$$|A_i| \geq m + 1 + (K - iN + i(i-1)/2)(m - 1)\ . \tag{4}$$

Property (ii): elements appearing in row $N_0 + i$ of A_{N_0+i}, $1 \leq i \leq N_1$ appear at least twice, and we claim that

$$|A_{N_0+i}| \geq m + 1 + (N_1 - 1 - i)(m - 1)\ . \tag{5}$$

Similar to the proof of type $\{5, 3\}$ we will show that these submatrices lead to a contradiction. The matrices satisfying Property (i) will contribute the top N_0 rows, while those satisfying Property (ii) will add the bottom N_1 rows. The double line in (3) shows the division in the $\{5, 3\}$ case.

We now prove the claim of Property (i) by induction on i. When creating A_1 from A_0, we want elements that appear in the first row to appear $N + 1$ times (in the first row). Let t_i be the number of elements repeating i times in the first row. We must delete at most $N(m-1)$ columns from A_0, for if it were the case

that $t_1 + t_2 + \ldots + t_N = Nm$, A would only have Nm columns. Assume (4) holds up to A_i. To ensure the elements appearing in row $i+1$ of A_{i+1} appear at least $N-i+1$ times in said row, we must delete $t_1+\ldots+t_{N-i} \leq (N-i)(m-1)$ columns of A_i. If $t_1 + \ldots + t_{N-i} = (N-i)m$, then A_i would have $(N-i)m$ columns, fewer than the number assumed in the inductive hypothesis. Then

$$\begin{aligned} |A_{i+1}| &\geq |A_i| - (N-i)(m-1) \\ &= m + 1 + (K - iN + i(i-1)/2 - N + i)(m-1) \\ &= m + 1 + (K - (i+1)N + i(i+1)/2)(m-1) \ , \end{aligned}$$

which proves the claim in Property (i).

Since Property (i) holds and $K = N_1 - 1 + N_0 N - N_0(N_0-1)/2$, note that

$$\begin{aligned} |A_{N_0}| &\geq m + 1 + (K - N_0 N - N_0(N_0-1)/2)(m-1) \\ &= m + 1 + (N_1 - 1)(m-1) \ . \end{aligned}$$

Now we consider the submatrices of Property (ii). When we create A_{N_0+1} we delete columns from A_{N_0} so that elements in row $N_0 + 1$ appear at least twice. The number of deleted columns does not exceed $m-1$, since we know A_{N_0} has more than m columns. Therefore $|A_{N_0+1}| \geq |A_{N_0}| - (m-1)$ as required. Assume Property (ii) holds up to i. We want A_{N_0+i+1} to have elements appearing at least twice in row $N_0 + i + 1$. Again, we delete no more than $m-1$ columns, since A_{N_0+i} has more than m columns by the inductive hypothesis. Then

$$|A_{N_0+i+1}| = |A_{N_0+i}| - (m-1) = m + 1 + (N_1 - 1 - (i+1))(m-1) \ ,$$

proving the claim of Property (ii).

In the very last submatrix,

$$\begin{aligned} |A_{N-1}| = |A_{N_0+N_1-1}| &\geq m + 1 + (N_1 - 1 - (N_1 - 1))(m-1) \\ &= m + 1 \ , \end{aligned}$$

which means the last row of A_{N-1} has a repeated element. We can permute the columns of A_{N-1} so that the two rightmost columns have the same element in row N. Since row $N-2$ of A_{N-2}, row $N-3$ of A_{N-3}, ..., row $N_0 + 1$ of A_{N_0+1} all have elements occurring twice, A_{N_0} has a submatrix isomorphic to

a	$*$	$*$	$*$	$*$	a
	$\ddots$				$\vdots$
$*$	$*$	x	$*$	$*$	x
$*$	$*$	$*$	y	$*$	y
$*$	$*$	$*$	$*$	z	z

,

in the widest case. This pattern covers N_1 rows, and we proceed assuming the widest case, when the above submatrix is $N_1 + 1$ columns wide.

By Property (i), elements that appear in row $N_0 - 1$ of A_{N_0-1}, appear at least $N - (N_0 - 1 - 2) = N_0 + N_1 - N_0 + 3 = N_1 + 3$ times. Then we have

b	b	$*$	$*$	$*$	$*$	$*$
$*$	a	$*$	$*$	$*$	$*$	a
$*$		$\ddots$				$\vdots$
$*$	$*$	$*$	x	$*$	$*$	x
$*$	$*$	$*$	$*$	y	$*$	y
$*$	$*$	$*$	$*$	$*$	z	z

.

Moving up row by row to A_0, Property (i) ensures that each row has elements repeating one more time than in the previous row. The staircase pattern is incrementally extended leftward by one up to A_0, which has a submatrix isomorphic to (numbered from right to left):

		$\ldots$	N_1	$\ldots$	4	3	2	1
c	c	$*$	$*$	$*$	$*$	$*$	$*$	$*$
$*$	$\ddots$	$\ddots$	$*$	$*$	$*$	$*$	$*$	$*$
$*$	$*$	b	b	$*$	$*$	$*$	$*$	$*$
$*$	$*$	$*$	a	$*$	$*$	$*$	$*$	a
$*$	$*$	$*$		$\ddots$				$\vdots$
$*$	$*$	$*$	$*$	$*$	x	$*$	$*$	x
$*$	$*$	$*$	$*$	$*$	$*$	y	$*$	y
$*$	$*$	$*$	$*$	$*$	$*$	$*$	z	z

.

Consider the following partition of the columns of A_0.

$$\mathcal{S}_d = \{1\} \cup \{N_1 + 2k : 0 \leq k, N_1 + 2k < N + 1\}$$
$$\mathcal{S}_w = \{2, \ldots, N_1\} \cup \{N_1 + 2k + 1 : 1 \leq k, N_1 + 2k + 1 < N + 1\}$$

It is easy to see that the first two subsets $\{1\}$ and $\{2, \ldots, N_1\}$ cannot be separated, this pattern formed the contradiction in the proof of Theorem 2.1. Likewise, the columns in the second subsets of $\mathcal{S}_w$ and $\mathcal{S}_d$ are the odd and even indexed columns of a staircase, which also cannot be separated. Therefore A is not an SHF of type $\{|\mathcal{S}_w|, |\mathcal{S}_d|\}$.

It remains to show that $|\mathcal{S}_w| = w$ and $|\mathcal{S}_d| = d$. First we show the second subsets are both of size $d - 1$. Note that the number of columns greater than or equal to N_1 is odd, since the staircase is a $d, d-1$ forbidden configuration. The column N_1 overlaps the two patterns, and belongs to the first part of $\mathcal{S}_w$ leaving $d - 1$ for each set. The first subset of $\mathcal{S}_w$ has size $N_1 - 1 = w - d + 1$, giving $|S_w| = w$. The first subset of $\mathcal{S}_d$ has size 1, so $|S_d| = d$.

Thus A is not an SHF of type $\{w, d\}$, contradicting our original assumption thereby proving the theorem.

The proof is still valid if the repeated elements in the bottom N_1 rows occur in fewer than N_1 columns. The columns $2, \ldots, N_1$ are in $\mathcal{S}_w$, so if the repeated

elements occupy fewer columns, say t columns, then A is not an SHF of type $\{w-t,d\}$ and therefore cannot be an SHF of type $\{w,d\}$. □

Extending to N rows (in the usual way) gives the following corollary.

Corollary 4.2. *If an $SHF(N;n,m,\{w,d\})$ exists, then*

$$n \leq (2dw-w)m^{\lceil \frac{N}{w+d-1} \rceil} - 2dw + w + 1.$$

Remark 4.3. *The bound above applies to all choices of w,d, however it is the strongest bound known only when $2 \leq d \leq w-2$. For the case $\{w,1\}$ see Theorem 2.1, for type $\{w,w\}$ see Theorem 1.4 and for type $\{w,w-1\}$ see [11]. For the PHF case, $\{1,1\}$, see [6].*

5 SHF of type $\{w_1, w_2, \ldots, w_t\}$

The bound for type $\{w,d\}$ gives the following bound for type $\{w_1, w_2, \ldots, w_t\}$.

Theorem 5.1. *Suppose A is an SHF$(N,n,m,\{w_1,\ldots,w_t\})$ where $w_1 \leq w_2 \leq \ldots \leq w_t$. Let $u = \sum_{i=1}^{t} w_i$. Then*

$$n \leq (2w_1-1)(u-w_1)m^{\lceil N/(u-1) \rceil} - w_1(2u-2w_1+1) + 1 \tag{6}$$

Proof. By an appeal to Theorem 1.2, A is also an SHF of type $\{w_1,\ldots,w_{t-1}+w_t\}$. Repeating this $t-2$ more times, we have that A is also an SHF of type $\{w_1, w_2+\ldots+w_t\}$. The upper bound on n follows directly from substituting $w=w_1$ and $d=u-w_1$ into Corollary 4.2. □

A remark similar to Remark 4.3 is appropriate here as better bounds may exist when $t=2$ and certainly do exist for PHF types. Additionally, a stronger bound is known for type $\{1,1,2\}$, see [5].

When $w_1 = 1$, it is easy to see that the bound in Theorem 5.1 is stronger than the one in Theorem 1.5. To compare the bound given above (6) to the one proven by Blackburn (1) for arbitrary types, we ignore the last term of (6) and focus on the coefficient of $m^{\lceil N/(u-1) \rceil}$. We fix u and define $f(w_1) = (2w_1-1)(u-w_1)$, a function which describes the coefficient in (6). The function f has its maximum at $w_1 = \frac{2u+1}{4}$, and it is an increasing function for $w_1 \leq \frac{2u+1}{4}$.

The coefficient of (1) is $\binom{u}{2}$. Since $w_1 \leq w_2 \leq \ldots \leq w_t$ and $t \geq 2$, we can be certain that $w_1 \leq \frac{u}{t} < \frac{2u+1}{4}$, and thus $f(w_1) \leq f(\frac{u}{t})$. We can now show that $f\left(\frac{u}{t}\right) \leq \binom{u}{2}$.

$$\begin{aligned} & f\left(\frac{u}{t}\right) \leq \binom{u}{2} \\ \iff \quad & \frac{2u^2}{t} - \frac{2u^2}{t^2} - u + \frac{u}{t} \leq \frac{u^2}{2} - \frac{u}{2} \\ \iff \quad & \frac{2u^2}{t} - \frac{2u^2}{t^2} \leq \frac{u^2}{2} + \frac{u}{2} - \frac{u}{t} \end{aligned}$$

Now the RS $\geq \frac{u^2}{2}$ (since $t \geq 2$) so it is sufficient to show that

$$\frac{2u^2}{t} - \frac{2u^2}{t^2} \leq \frac{u^2}{2} .$$

This is equivalent to

$$\frac{2(t-1)}{t^2} \leq \frac{1}{2} .$$

We can see the effect of increasing t, as follows. The coefficient in (6) when $w_1 = \frac{u}{t}$ is approximately

$$\left(\frac{2u}{t}\right) \frac{u(t-1)}{t} = \frac{2u^2(t-1)}{t^2} .$$

This equals $\frac{u^2}{2}$ when $t = 2$ and decreases as t increases. The coefficient of $m^{\lceil \frac{N}{u-1} \rceil}$ in (1) is approximately $\frac{u^2}{2}$, independent of t. In the case $t = 2$ when the bounds are equal, Theorem 4.1 gives the strongest bound. When $t \geq 3$, the bound of Theorem 5.1 is stronger than the bound of Theorem 1.5 for all sets $\{w_1, \ldots, w_t\}$.

6 Conclusion

We have shown a necessary condition for SHF$(N; n, m, \{w_1, \ldots, w_t\})$. Some questions that remain open in this area are finding constructions and sufficient conditions. With respect to explicit constructions, some methods for $t = 2$ are given by Liu and Shen [1], Stinson, Wei and Zhu [3] and Tonien and Safavi-Naini [4]. It remains an open problem to explicitly construct good SHF of type $\{w_1, \ldots, w_t\}$ when $t \geq 3$.

References

[1] L. Liu and H. Shen. Explicit constructions of separating hash families from algebraic curves over finite fields. *Designs, Codes and Cryptography*, **41** (2006), 221–233.

[2] Y. L. Sagalovich. Separating systems. *Problems of Information Transmission*, **30** (1994), 105–123.

[3] D. R. Stinson, R. Wei and L. Zhu New constructions for perfect hash families and related structures using combinatorial designs and codes *Journal of Combinatorial Designs*, **8** (2000), 189–200.

[4] D. Tonien and R. Safavi-Naini Recursive constructions of secure codes and hash families using difference fucntion families *Journal of Combinatorial Theory Series A*, **113** (2006), 664–674.

[5] D. R. Stinson, R. Wei and K. Chen. On generalized separating hash families. *Preprint.* Available online: `http://www.cacr.math.uwaterloo.ca/~dstinson`

[6] S. R. Blackburn and P. R. Wild. Optimal linear perfect hash families. *J. Comb. Theory Ser. A*, **83** (1998), 233–250.

[7] S. R. Blackburn. An upper bound on the size of a code with the k-identifiable parent property. *Journal of Combinatorial Theory Series A*, **102** (2003), 179–185.

[8] S. R. Blackburn. Frameproof codes. *SIAM Journal on Discrete Mathematics*, **3** (2003), 499–510.

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

[10] A. Barg, G. R. Blakeley and G. Kabatiansky. Digital fingerprinting codes: problem statements, constructions and identification of traitors. *IEEE Transactions on Information Theory*, **49** (2003), 852–865.

[11] D. R. Stinson and G. M. Zaverucha. Some improved bounds for secure frameproof codes and related separating hash families, *Preprint.* (2007) Available online: `http://www.cacr.math.uwaterloo.ca/~dstinson`

[12] S. R. Blackburn. A note on separating hash families. *Preprint.* (2007).

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