Mathematical Problems of Computer Science 21, 2000, 104–115.

New Upper Bound on the Cardinality of a k-separated Set or Perfect Hash Family and a Near Optimal Construction for It

Samvel S. Martirosyan, Sosina S. Martirosyan

Institute for Informatics and Automation Problems of NAS RA and YSU

Abstract

In this paper a rather simple method to obtain upper bound on the cardinality of so called k-separated (kS) sets or perfect Hash families is presented. The obtained bound is better than similar Fredman-Komlòs [4] and Körner-Marton [6] bounds almost everywhere. For some cases the exact value on the cardinality of a kS set is given. Besides two constructions of kS sets are given. The first construction is recursive in nature, where for arbitrary q and k $(q \geq k)$, n is practically of order $\log N$ and in this sense is near to optimal. The second construction method allows to give good constructions of $3S$ sets for small values of n.

1 Introduction

Let D be some subset of $\{1, 2, \ldots, q\}^n$ represented in the form of a $|D| \times n$ matrix D.

Definition 1. The set D is called kS $(k \leq q)$, if any k -rowed submatrix of D includes at least one column all elements of which are distinct.

Let us denote the maximal cardinality of a kS set by

$$N(n, q, k) = \max_{D} |D|.$$

The problem on obtaining good upper and lower bounds on $N(n, q, k)$ or development of good constructions for kS sets has recently grown into an urgent one. These sets are also known in literature under the name of *perfect Hash Families.*

Definition 2. (see [7]).

a) A function $h : \{1, 2, \ldots, N\} \to \{1, 2, \ldots, q\}$ is a perfect *hash* function for $S \subseteq \{1, 2, \ldots, N\}$, if $h(x) \neq h(y)$ for all $x, y \in S$, $x \neq y$.

b) A family H of functions $h : \{1, 2, \ldots, N\} \to \{1, 2, \ldots, q\}$ is called (N, q, k) perfect if for every $S \subseteq \{1, 2, \ldots, N\}$, $|S| = k$, there exists a $h \in H$ such that h is perfect for S.

If we denote $n = |H|$, then we have $\max_{|H|=n} N = N(n, q, k)$.

Earlier results obtained on upper and lower bounds of these sets. Fredman and Komlós in [4] have obtained upper and lower bounds on $N(n, q, k)$ expressed in the following form

(see [6]):

$$\frac{1}{k-1}\log\frac{1}{1-\frac{q^k}{q^k}} \underset{\sim}{\leq} \frac{1}{n}\log N(n,q,k) \underset{\sim}{\leq} \frac{q^{k-1}}{q^{k-1}}\log(q-k+2) \tag{1}$$

where $q^k = \prod_{j=0}^{k-1}(q-j)$, and $A \underset{\sim}{\leq} B$ denotes that $A \leq (1+0(1))B$, where $0(1)$ tends to zero when n tends to infinite.

In [5], [6] Körner and Marton using a very complicated proof technique, improved the upper bound on $N(n,q,k)$ as follows:

$$\frac{1}{n}\log N(n,q,k) \underset{\sim}{\leq} \min_{0\leq j\leq k-2}\frac{q^{j+1}}{q^{j+1}}\log\frac{q-j}{k-j-1} \tag{2}$$

This bound matches with the one given in (1) for $j = k-2$.

Both upper bounds in (1) and (2) are of asymptotic nature and practically provide no means to obtain the exact value of n.

Earlier results obtained on constructions. The known upper and lower bounds {see for example bound (1)) theoretically state that for arbitrary q and k $(q \geq k)$ there exists a kS set such that the length n and the cardinality N are connected by the relation

$$n = \Theta(\log N).$$

Some constructions of kS sets expressed by error-correcting codes and balanced incomplete block designs(BIBD) are presented in papers [1] and [3], respectively. However the paper [2] convincingly shows that for both constructions the length n strongly depends on q and k and therefore they don't allow to construct kS sets for large n for the given q and k.

In paper [2] two recursive constructions of kS sets, where n is a polynomial function of $\log N$ (for fixed q and k) are given. More precisely

$$n = \Theta\left((\log N)^{\log\binom{k}{2}+1}\right).$$

In this paper using a rather simple method, we obtain the upper bound on $N(n,q,k)$ which is non-asymptotic and is better than similar Fredman-Komlós and Körner-Marton bounds almost everywhere. For some cases of interest the exact value of $N(n,q,k)$is given.

Besides two constructions of kS sets are given. The first one- recursive in nature,for arbitrary q and k $(q \geq k)$, n is practically of order $\log N$ and in this sense is near to optimal. The second method allows to give good constructions for $3S$ sets for small values of n.

2 Upper bounds

To obtain the upper bounds we need the following two Lemmas:

Lemma 1. *Let D be a $N(n,q,k) \times n$ kS matrix. Then for an arbitrary q the following relations hold:*

(i) if $k \geq 2n$, then

$$N(n,q,k) = q.$$

(ii) if $k = 2n-1$, *then*

$$N(n, q, k) \le \lfloor \tfrac{k+1}{k-1}(q-1) \rfloor, \quad \text{if} \quad n-1 \text{ does not divide } q-1$$
$$N(n, q, k) = \tfrac{k+1}{k-1}(q-1), \quad \text{if} \quad n-1 \text{ divides } q-1.$$

(iii) if $k = 2n-2$, *then*

$$N(n, q, k) \le \begin{cases} \left\lfloor \frac{k-2}{k-4}(q-1) \right\rfloor & \text{if } n > 3 \\ 3q-6, & \text{if } n = 3. \end{cases}$$

Proof. For our convenience, we call an element of the alphabet that occurs more than once in any column of the matrix D as a *special element* and denote $N(n, q, k)$by N.

Proof of (i). Let $k \ge 2n$. Suppose that $N > q$. Then, evidently, each column in the matrix D contains at least a single special element. Choose a pair of such elements from each column and mark out those rows in matrix D where the chosen elements stand. Clearly, the number of such rows in the matrix is not more than $2n$. Then any k- rowed submatrix of D with marked out rows will contain no column whose all elements are distinct. Thus we obtain that the matrix D is not kS. This is a contradiction which proves (i).

Proof of (ii). Let $k = 2n-1$. On one hand, it is clear that the number of special elements in each column of the matrix is not less than $N-q+1$.(This number is equal to $N-q+1$, if in any column of the matrix there exists only a single element of such kind). On the other hand, no row of the matrix may contain more than a single special element, otherwise there will be $2n-1$ rows of the matrix D which do not contain a column whose all elements are distinct.

Hence, we may write

$$(N-q+1)n \le N$$

or

$$N \le \lfloor \frac{n(q-1)}{n-1} \rfloor = \lfloor \frac{k+1}{k-1}(q-1) \rfloor. \tag{3}$$

On the other hand the construction given below shows that if $k = 2n-1$,and $(n-1)$ divides $(q-1)$, then the bound (3) is achievable.

Construction. Let $k = 2n-1$ and $(n-1)$ divides $(q-1)$. Then we'll take any vector of length $\frac{n(q-1)}{n-1}$ with a single special element in its $(i-1)\frac{q-1}{n-1}$ to $i\frac{q-1}{n-1}$ positions as the i-th $(i = 1, \ldots, n)$ column of the matrix. Naturally, this matrix will be $k = 2n-1$- separated.

Example. $n = 5$, $k = 9$, $q = 13$. Since $4 \mid 12$ then we have a 15×5 matrix which is a 9-separated optimal set:

$$D = \begin{bmatrix} 1 & 2 & 2 & 2 & 2 \\ 1 & 3 & 3 & 3 & 3 \\ 1 & 4 & 4 & 4 & 4 \\ 2 & 1 & 5 & 5 & 5 \\ 3 & 1 & 6 & 6 & 6 \\ 4 & 1 & 7 & 7 & 7 \\ 5 & 5 & 1 & 8 & 8 \\ 6 & 6 & 1 & 9 & 9 \\ 7 & 7 & 1 & 10 & 10 \\ 8 & 8 & 8 & 1 & 11 \\ 9 & 9 & 9 & 1 & 12 \\ 10 & 10 & 10 & 1 & 13 \\ 11 & 11 & 11 & 11 & 1 \\ 12 & 12 & 12 & 12 & 1 \\ 13 & 13 & 13 & 13 & 1 \end{bmatrix}.$$

Proof of (iii). Let $k = 2n - 2$, $n > 3$.

First, suppose that for any two columns of matrix D, there exist no row in the matrix which includes a special element chosen from both columns. Then we have the following inequality

$$n(N - q + 1) \leq N$$

or

$$N \leq \lfloor \frac{n(q-1)}{n-1} \rfloor = \lfloor \frac{k+2}{k}(q-1) \rfloor \tag{4}$$

Now suppose that in the matrix D there exist two columns such that they both simultaneously have a special element at least on one row. Then there should not be any other two columns among the remaining $n-2$ columns in the matrix, which have a special element on any row, otherwise the matrix D will not be $k = 2n - 2$ -separated. Hence, we have

$$(n-2)(N - q + 1) \leq N$$

or

$$N \leq \lfloor \frac{n-2}{n-3}(q-1) \rfloor = \lfloor \frac{k-2}{k-4}(q-1) \rfloor \tag{5}$$

Comparing (4) with (5), we obtain the first case of (iii).

Now consider the case for $n = 3$, $k = 4$.

First, suppose that there exist two columns in the matrix D which simultaneously have no special element on any row of the these columns. Then we have the inequality

$$2(N - q + 1) \leq N$$

or

$$N \leq 2(q-1) \tag{6}$$

Now let two columns in each of three pairs of columns simultaneously have a special element on any row of the matrix. In this case each column has at least two distinct special elements and there are no more than two such elements on each row, otherwise the matrix will be no $4S$.

Hence, since the number of special elements is no less than $N-q+2$ in this case, then we have $3(N-q+2) \leq 2N$ or

$$N \leq 3q-6 \tag{7}$$

From (6) and (7) follows the second case in (iii). This fully completes the proof of Lemma 1. ■

(Note that for large k, in a manner exactly analogous to the one given in Lemma 1, one may obtain bounds better than the presented ones).

Lemma 2. *For all values of parameters n, q, k $(k \leq q)$, the following inequality holds:*

$$N(n, q, k) \leq \lfloor N(n-1, q, k) \frac{q}{k-1} \rfloor .$$

Proof. Let D be a kS set.

Consider any of the columns in matrix D. For definiteness, let it be the 1-st column, and let x_i, $1 \leq i \leq q$, be the number of i symbol in that column. Then $\sum_{i=1}^{q} x_i = N(n, q, k)$. Choose $k-1$ greatest numbers of $x_1, x_2, \ldots, x_q$. Let they be the first $k-1$ $(x_1, x_2, \ldots, x_{k-1})$ ones. Assume $\sum_{i=1}^{k-1} x_i = M$. Then, on one hand, we have

$$M \geq \frac{N(n, q, k)}{q} \cdot (k-1), \tag{8}$$

since $N(n, q, k)/q$ is the mean multiplicity of the element in that column.

On the other hand, consider a $M \times (n-1)$ matrix D' which includes the first M rows of matrix D without its first column. This matrix is kS. Hence, we have

$$M \leq N(n-1, q, k). \tag{9}$$

From expressions (8) and (9) we obtain

$$\frac{N(n, q, k)}{q} \cdot (k-1) \leq N(n-1, q, k)$$

which proves Lemma 2. ■

From Lemmas 1 and 2 follows:

Theorem 1. *The upper bound on $N(n, q, k)$ is given by:*

$$N(n, q, k) \leq \underbrace{\lfloor \lfloor \lfloor A \frac{q}{k-1} \rfloor \frac{q}{k-1} \rfloor \times \cdots \times \frac{q}{k-1} \rfloor}_{n-n_1} \tag{10}$$

where

$$A = \begin{cases} \left\lfloor \frac{k+1}{k-1}(q-1) \right\rfloor & \text{for } k = 2n_1 - 1 \\ \left\lfloor \frac{k-2}{k-4}(q-1) \right\rfloor & \text{for } k = 2n_1 - 2, \quad n_1 > 3 \\ 3q-6 & \text{for } k = 4 \qquad\quad n_1 = 3. \end{cases}$$

■

Comparison of upper bounds (1), (2) and (10)

First, we compare the bounds (1) and (2). To do this we state the following lemma:

Lemma 3. *For any fixed number k, there exists a number $q_0(k)$ such that for all $q > q_0(k)$, the minimum of the right side in Körner-Marton bound*

$$\frac{1}{n}\log N(n, q, k) \underset{\sim}{\leq} \min_{0 \leq j \leq k-2} \frac{q^{j+1}}{q^{j+1}} \log \frac{q-j}{k-j-1} \tag{11}$$

is achieved for $j = 0$.

Proof. Denote by

$$A_j(q, k) = \frac{q^{j+1}}{q^{j+1}} \log \frac{q-j}{k-j-1}.$$

Now estimate the difference of $A_{j+1}(q, k) - A_j(q, k)$, $0 \leq j \leq k-3$.

$$A_{j+1}(q, k) - A_j(q, k) = \frac{q^{j+2}}{q^{j+2}} \log \frac{q-j-1}{k-j-2} - \frac{q^{j+1}}{q^{j+1}} \log \frac{q-j}{k-j-1} =$$

$$= \frac{q^{j+1}}{q^{j+2}} \left[(q-j-1) \log \frac{q-j-1}{k-j-2} - q \log \frac{q-j}{k-j-1}\right] =$$

$$= \frac{q^{j+1}}{q^{j+2}} \left[\log \left[\frac{(q-j-1)(k-j-1)}{(q-j)(k-j-2)}\right]^q - \log \left(\frac{q-j-1}{k-j-2}\right)^{j+1}\right].$$

Since for $q \geq k$ $\frac{q-j-1}{q-j} \cdot \frac{k-j-1}{k-j-2} > 1$, then $\left[\frac{(q-j-1)(k-j-1)}{(q-j)(k-j-2)}\right]^q$ increases faster with q than $\left(\frac{q-j-1}{k-j-2}\right)^{j+1}$. Then it is clear that there exists a number $q_0(k, j)$ such that for $q > q_0(k, j)$ the later expression is positive.

Take

$$q_0(k) = \max_{0 \leq j \leq k-3} q_0(k, j).$$

Then for all $q > q_0(k)$ $A_j(q, k)$ increases monotonely with j, which proves Lemma 3. ∎

Lemma 3 shows that except for a finite number of values of q the Körner-Marton bound is better than Fredman-Komlós bound.

The bound (2) is also given in [6] in the following form:

$$\frac{1}{n}\log \frac{N(n, q, k) - j}{k-j-1} \leq \frac{q^{j+1}}{q^{j+1}} \log \frac{q-j}{k-j-1}, \quad \text{for any } j = 1, 2, \ldots, k-2.$$

In view of all facts mentioned above, for $q > q_0(k)$ we have the inequality

$$N(n, q, k) \leq (k-1)\left(\frac{q}{k-1}\right)^n$$

Comparing this bound with (10) we come to the following:

Conclusion. For a fixed k there exists a $q_0(k)$ such that for $q > q_0(k)$ Körner-Marton bound is better than Fredman-Komlós bound. In the cases when Körner-Marton bound is better than Fredman-Komlós bound, then bound (10) is better than both these bounds.

Illustrate this by an example.

Example. Let $q = 9$, $k = 5$, then the bounds (for $n > 3$)

bound (1) $\quad N(n, 9, 5) \underset{\sim}{\leq} (2.2837)^n$

bound (2) $\quad N(n, 9, 5) \leq 4 \cdot (2.25)^n$

bound (10) $\quad N(n, 9, 5) \leq \underbrace{\lfloor\lfloor\lfloor 1.1415 \cdot 2.25\rfloor 2.25\rfloor \times \ldots \times 2.25\rfloor}_{n}.$

3 Constructions

First Construction. To describe the algorithm of the first construction we use the following two Theorems.

Theorem 2 [1]. *Let $V \subset \{0,1,\ldots,q-1\}^n$ be a code with minimum Hamming distance d. If*

$$d > \left(1 - \frac{1}{\binom{k}{2}}\right) n,$$

then V is also kS set.

Proof. Components of any two vectors $\overline{a}$ and $\overline{b}$, $\overline{a}, \overline{b} \in V$ differ at least in d positions and components of any third vector $\overline{c}$, $\overline{c} \in V$, $\overline{c} \neq \overline{a}, \overline{b}$ coincide with the components of each of two vectors $\overline{a}$ and $\overline{b}$ no more than in $n-d$ positions. Hence if

$$d > 2(n-d) \text{ or } d > \frac{2}{3}n,$$

then the set V is $3S$. Moreover in a matrix including any three vectors of V there exist at least $3d-2n$ columns whose all elements are distinct, and therefore if the following inequality also holds

$$3d - 2n > 3(n-d) \text{ or } d > \frac{5}{6}n,$$

then in the matrix of any four vectors of V there exist at least $6d-5n$ columns whose all elements are distinct. Thus V is $4S$.

Suppose, proceeding in this manner we come to the following:

If

$$d > \frac{a_{k-1}}{b_{k-1}} n,$$

where a_{k-1} and b_{k-1} are natural numbers, then V is a $(k-1)S$ set and any matrix of $(k-1)$ vectors of V has at least $b_{k-1}d - a_{k-1}n$ columns with all distinct elements. And if the inequality also holds

$$b_{k-1}d - a_{k-1}n > (k-1)(n-d)$$

or

$$d > \frac{a_{k-1}+k-1}{b_{k-1}+k-1} n = \frac{a_k}{b_k} n$$

then V is a kS set.

By taking $a_3 = 2$ and $b_3 = 3$ in the following recurrent relations:

$$a_k = a_{k-1} + k - 1$$
$$b_k = b_{k-1} + k - 1$$

we have

$$a_k = \frac{k(k-1)}{2} - 1 = \binom{k}{2} - 1$$
$$b_k = \frac{k(k-1)}{2} = \binom{k}{2},$$

which proves Theorem 2.

■

[1] This theorem was already proved by the authors when they knew that using some other technique, N.Alon have given the proof of this Theorem earlier (see [1]).

The following two corollaries are easily obtained from this Theorem.

For convenience we denote kS sets by $PHF(n, N, q, k)$ following the notations accepted by the authors in paper [2].

Corollary 1. Let V be q-ary MDR code (N, K, D), $K + D = N + 1$. If $K = \lfloor \frac{N}{\binom{k}{2}} \rfloor$, then V is $PHF\left(N, q^k, q, k\right)$.

Corollary 2. Let V be extended Reed-Solomon(RS) code over $GF(p^m)$, $(N = p^m, K, D = N - K + 1)$, where $p \geq \binom{k}{2}$. If $K = p^{m-1}$, then V is $PHF\left(p^m, p^{mp^{m-1}}, p^m, k\right)$.

Theorem 3. *If there exist* $PHF(n_0, N_0, q_0, k)$ and $PHF(n_1, N_1, q_1, k)$, *where* $q_0 < q_1 \leq N_0$, *then there also exists* $PHF(n_0 n_1, N_1, q_0, k)$.

Proof. Put each symbol from the set $\{1, 2, \ldots, q_1\}$ in correpondence with any distinct row vectors of the set $PHF(n_0, N_0, q_0, k)$. Then substitute each symbol of $\{1, 2, \ldots, q_1\}$ in the matrix $PHF(n_1, N_1, q_1, k)$ by corresponding vector of length n_0. We obtain $N_1 \times n_0 n_1$ matrix consisted of symbols $\{1, 2, \ldots, q_0\}$. This matrix is $PHF(n_0 n_1, N_1, q_0, k)$. Indeed, by the condition of the Theorem in any k rows of matrix $PHF(n_1, N_1, q_1, k)$there exists a column with k distinct symbols from $\{1, 2, \ldots, q_1\}$. A $N_1 \times n_0$ submatrix in the matrix $PHF(n_0 n_1, N_1, q_0, k)$ corresponds to this column on the same k rows of which stand k vectors from matrix $PHF(n_0, N_0, q_0, k)$. Therefore, by the condition of the theorem this submatrix includes a column which contains k distinct symbols from $\{1, 2, \ldots, q_0\}$.

■

Construction algorithm.

Let q, k $(q \geq k)$ be arbitrary natural numbers. The construction algorithm is implemented step by step.

Step 0. Using some method construct $PHF(n_0, N_0 = p^i, q, k)$for arbitrary $i \geq 2$ and prime $p \geq \binom{k}{2}$.

Step 1.

We have $PHF(n_0, N_0 = p^i, q, k)$.

Following Corollary 2 construct $PHF\left(n_1' = p^i, N_1 = p^{ip^{i-1}}, p^i, k\right)$.

By Theorem 3 we obtain $PHF\left(n_1 = n_0 n_1', N_1 = p^{ip^{i-1}}, q, k\right)$.

Thus we have

$$n_1 = \frac{n_0}{i \log p} \cdot p \log N_1$$

$\vdots$

Step l.

From the step $l - 1$, we have $PHF(n_{l-1}, N_{l-1}, q, k)$.

Using Corollary 2 construct $PHF\left(n_l' = N_{l-1}, N_{l-1}^{\frac{N_{l-1}}{p}}, N_{l-1}, k\right)$.

By Theorem 3 we obtain $PHF\left(n_l = n_{l-1} n_l', N_l = N_{l-1}^{\frac{N_{l-1}}{p}}, q, k\right)$.

Applying induction method we get the following relation between n_l and N_l

$$n_l = \frac{n_0}{i \log p} \cdot p^l \log N_l.$$

Bring the latter relation into the form

$$n = \Theta\left(\varphi\left(\log N\right)\right).$$

From the recurrent relation

$$N_l = N_{l-1}^{\frac{N_{l-1}}{p}} \quad \text{and} \quad N_0 = p^i,$$

we obtain

$$N_l = p^{\alpha_{l-1} p^{\alpha_{l-1} p^{\alpha_{l-2} p^{\cdots^{\alpha_1 p^{\alpha_0 p^{i-1}}}}}}}, \tag{12}$$

where $\alpha_0 = i$, for $1 \le j \le l-1$ $\alpha_j = \frac{N_{j-1}}{p}$.
Taking $\alpha_j = 1$ for all $1 \le j \le l-1$ in (12), we have

$$N_l > p^{p^{\cdots^{p^{i-1}}}}.$$

From which, if $1 \le l \le i-1$, since $p^i < ip^{i-1}$, then

$$n_l = \Theta\left(\underbrace{\log_p \ldots \log_p}_{l-1} N_l \times \log N_l\right),$$

and if $i-1 < l \le ip^{i-1}$, then

$$n_l = \Theta\left(\underbrace{\log_p \log_p \ldots \log_p}_{l-2} N_l \times \log N_l\right),$$

etc.

In general, we have

$$n = \Theta\left(\underbrace{\log_p \log_p \ldots \log_p}_{s} N \times \log N\right),$$

where when $n \to \infty$, $s \to \infty$.

Evidently, the first multiplier in the parentheses is far much less than $\log N$, and therefore we may say that the order of n is practically near to asymptotically optimal value $\Theta\left(\log N\right)$.

Example. $k = 3, q = 3$.

The constructions described in [2] for these parameters give the following $3S$ sets:

(a) $PHF\left(3 \times 4^j, 5^{2^j}, 3, 3\right)$, $\quad n \approx 0.556\left(\log N\right)^2$;

(b) $PHF\left(2j\left(j-1\right), 3^j, 3, 3\right)$, $\quad n \approx 0.796\left(\log N\right)^2$;

(c) our construction: We'll use RS codes with the parameters $(q-1, k, d)$, where $q = 3^i$, $k = 3^{i-1}$ and $d = 2 \cdot 3^{i-1} > \frac{2}{3}\left(3^i - 1\right)$ as MDR of corollary 1.

Step 0. Take $PHF(4,9,3,3)$-so called 'tetra' code.

Step 1. We have $PHF(4,3^2,3,3)$.
By Corollary 1 construct $PHF(8,3^6,9,3)$.
By Theorem 3 we obtain $PHF(32,3^6,3,3)$.

Step 2. We have $PHF(32,3^2,3,3)$.
By Corollary 1 construct $PHF\left(3^6-1,3^{6\cdot3^5},3^6,3\right)$.
By Theorem 3 we obtain $PHF(23296,3^{1458},3,3)$.

Using the table below we may compare the constructions (a), (b) with (c) for some values of n and N.

Constructions	n	N
(a), $j=2$	48	625
(b), $j=6$	60	729
(c), step 1	32	729
(a), $j=7$	49125	$5^{128}\approx 3^{185}$
(b), $j=1458$	4248612	3^{1458}
(c), step 2	23296	3^{1458}

Construction 2. (For $q=3$, $k=3$).
Construction algorithm.
The construction algorithm is implemented step by step. In the j-th step we construct

$$PHF(j^2,3^j,3,3).$$

Denote the corresponding matrix by M_j.
Step 0: Take the following 3 matrices

$$a=\begin{pmatrix}0&0&0\\1&1&1\\2&2&2\end{pmatrix},\quad b=\begin{pmatrix}0&1&2\\1&2&0\\2&0&1\end{pmatrix},\quad c=\begin{pmatrix}0&2&1\\1&0&2\\2&1&0\end{pmatrix}.$$

Remark. It is easily verified that the rows of any two of these three matrices taken together are $PHF(3,6,3,3)$.
Denote by L_j the $3^j\times j$ matrix consisted of all vectors of length j from the alphabet $\{a,b,c\}$.
Step 1. For $j=1$

$$M_1=\begin{pmatrix}0\\1\\2\end{pmatrix}$$

we have $PHF(1,3,3,3)$.
Step 2. For $j=2$ construct the matrix

$$M_2'=(M_1L_1)=\begin{pmatrix}0&a\\1&b\\2&c\end{pmatrix}.$$

Tripling the rows in matrix M_1 and substituting the symbols a, b, c by corresponding matrices, we obtain 9×4 size matrix

$$M_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ \\ 1 & 0 & 1 & 2 \\ 1 & 1 & 2 & 0 \\ 1 & 2 & 0 & 1 \\ \\ 2 & 0 & 2 & 1 \\ 2 & 1 & 0 & 2 \\ 2 & 2 & 1 & 0 \end{bmatrix}$$

and obtain $PHF(2^2, 3^2, 3, 3)$ – so called tetra code.

Step j.

Lemma 4. *Let the $3^{j-1} \times (j-1)$matrix M_{j-1} be $PHF\left((j-1)^2, 3^{j-1}, 3, 3\right)$. Then the matrix*

$$M_j'' = \begin{pmatrix} M_{j-1}^* & L_{j-1}^* \end{pmatrix}$$

of size $3^j \times \left((j-1)^2 + 3(j-1)\right)$, where M_{j-1}^* is obtained from M_{j-1} by tripling the rows in it and L_{j-1}^* is obtained from L_{j-1} by substituting of symbols a, b, c by corresponding matrices, is $PHF\left((j-1)^2 + 3(j-1), 3^j, 3, 3\right)$.

Proof. Let $\overline{x} = \overline{x}_m\overline{x}_l$, $\overline{y} = \overline{y}_m\overline{y}_l$, $\overline{z} = \overline{z}_m\overline{z}_l$ be three distinct rows in matrix M_j^*. Then

1. Let $\overline{x}_m \neq \overline{y}_m \neq \overline{z}_m$. Then the proof follows immediately from the condition of the Lemma.

2. Let $\overline{x}_m = \overline{y}_m = \overline{z}_m$. Then any column of matrix $\begin{pmatrix} \overline{x}_l \\ \overline{y}_l \\ \overline{z}_l \end{pmatrix}$ is consisted of distinct symbols.

3. Let any two vectors of three vectors $\overline{x}_m$, $\overline{y}_m$, $\overline{z}_m$ coincide, but the third vector does not, let namely $\overline{x}_m = \overline{y}_m \neq \overline{z}_m$. Then in matrix

$$\begin{pmatrix} \overline{x}_l \\ \overline{y}_l \\ \overline{z}_l \end{pmatrix}$$

exists a column whose all elements are distinct, which follows from the Remark.

This completes the proof of Lemma 3. ■

It is easily seen that the column of the form $(012012\ldots012)^T$ occurs $(j-1)$ times in the matrix M_j''. Leaving only one of these columns in the matrix , we obtain the matrix M_j consisted of l^2 columns.

References

[1] N. Alon. *Explicit Construction of Exponential Sized Families of k-independent Sets.* Discrete Mathematics 58 (1986) 191-193.

[2] M. Ataci, S. S Magliveras, D. R. Stinson and W. D. Wei. *Some Resursive Constructions for Perfect Hash Families.* J. Combinatorial Design 4 (1996), pp. 353-363.

[3] E. F. Brickell. *A problem in Broadcast Encryption.* Presented at the Fifth vermont Summer Worshop on Combinatorics and Graph Theory, June 1991.

[4] M. Fredman and J. Komlós. *On the Size of Separating Systems and Perfect Hash Functions.* SLAM J. Algebraic Discr. Meth. 5 (1984) 61-*xpli*18.

[5] J. Körner. *Fredman-Komlós Bounds and Information Theory.* SLAM J. Algebraic Discr. Meth. 7 (1986) 560–570.

[6] J. Körner and K. Marton. *New Bounds for Perfect Hashing via Information Theory.* Europ. J. Combinatorics. 9 (1988) 523–530.

[7] K. Mehlhorn. *Data Structures and Algorithms.* , vol. 1. Springer-Verlag. 1984.