S©MR ISSN 1813-3304

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports
http://semr.math.nsc.ru

Том 15, стр. 561–569 (2018) УДК 519.143, 519.728
DOI 10.17377/semi.2018.15.045 MSC 05B15, 94C30

PARTIAL COVERING ARRAYS FOR DATA HIDING AND QUANTIZATION

V.N. POTAPOV

Abstract. We consider the problem of finding a set (partial covering array) S of vertices of the Boolean n-cube having cardinality 2^{n-k} and intersecting with maximum number of k-dimensional faces. We prove that the ratio between the numbers of the k-faces containing elements of S to k-faces is less than $1 - \frac{1+o(1)}{2^{k+1}}$ as $n \to \infty$. The solution of the problem in the class of linear codes is found. Connections between this problem, cryptography and an efficiency of quantization are discussed.

Keywords: linear code, covering array, data hiding, wiretap channel, quantization, wet paper stegoscheme.

1. Introduction

Let F_2^n be the set of binary vectors of length n (hypercube). A k-dimensional face (k-face) of hypercube is obtained by fixing any $n-k$ coordinates of the vectors in F_2^n. We consider some problems relating to information transmission. The first problem is the message transmission over wiretap channel [6]. Consider the following situation. An adversary can intercept $n-k$ bits (in random positions) of an n-bit message. The encoder is to be designed to minimize the adversary's information about the initial data. A general approach for solving this problem is to split hypercube into $C_1, \dots, C_{2^k}$ sets, for example, by syndromes of some linear code, and to encode k-bit data x by a random n-bit word from C_x. Let Γ be k-face defined by intercepted $n-k$ bits. Then the adversary is forced to choose between all x such that $C_x \cap \Gamma \neq \varnothing$. The encoder needs a partition such that each k-face of

Potapov, V.N., Partial covering arrays for data hiding and quantization.

The work was supported by the program of fundamental scientific researches of the SB RAS №I.5.1., project №0314-2016-0017 and by the program of fundamental scientific researches of Presidium of the Russian Academy of Sciences №0314-2015-0011.

Received August, 5, 2017, published May, 11, 2018.

the hypercube intersects with as many as possible elements of partition. In other words each C_x must intersect as many k-faces as possible.

One of a well-known stegoscheme is based on coding theory. Encoder changes one or more bits of the initial message in order to the resulting word w has a special syndrome x, i. e. $u \in C_x$. This syndrome x is a hiding message. It is assumed that the changes of initial message are not perceptible. However, data obtained by modern methods of coding images contain control bits of different kind that cannot be changed. So called wet paper stegoscheme divides the coordinates into wet coordinates that can be used for hiding information and dry coordinates that cannot be altered [5]. However alternation of different wet coordinates corresponds to hugely different effects. Places of the least significant bits depend of image. Consider this issue in detail.

The efficient quantization of real data is an important stage for lossy compression of images or speechs. From the nature of the things a part of data values is on the edges of the quantization intervals. The last bit of such value is the least significant one for the quality of quantization. Thus useful stegoscheme should provide possibility of alternation of different bits for embedding the same message, i.e. for providing the syndrome x. We conclude that the set of words C_x generated by stegoscheme must intersect as many k-faces as possible where k is the maximum number of alternating bits. A special method for choosing this least significant bits is used for data hiding in image and video [8].

It is possible to utilize this redundancy for data compression [1]. Consider an n-tuple consisted of the last bits of quantized values. Suppose that each n-tuple contains k' negligible bits. Let C be some code (array of codewords) with cardinality $2^{n-k'}$ and let for each n-tuple there exists a codeword such that this n-tuple and the codeword differ only in negligible bits. We will transmit the codeword (rather its number) instead of the initial n-tuple. So, we will truncate k' bits of the n-bit message.

Consider a mathematical formulation of the problem. We would like to construct a minimal code such that for any binary n-tuple v and for each set of k positions (that contains the least significant bits) there exists a codeword u such that v and u can differ from each other only in these k positions. I.e., the k-face defined by these positions contains a codeword. We need either to find such code (covering array) with the least cardinality or to fix a cardinality of the code and to maximize the number of the fit positions. Note that this problem is different from the problem of construction of covering codes. A code C is called k-covering code if for every n-tuple there exists a codeword which is different in certain k positions. But we going to maximize the number of the fit k-sets.

A perfect solution for the problems written above would be a set of codewords containing only one element of each k-face of F_2^n. Such sets have cardinalities 2^{n-k} and are called MDS codes. For the q-ary hypercubes (q is a prime power), there are MDS codes with several code distances. If the cardinality of an MDS code is q^{n-k} then the code distance is equal to $k+1$. But in the Boolean hypercube, there exist only two nonequivalent MDS codes: the parity check code ($k = 1$) and the pair of antipodal vectors ($k = n-1$), for example, $\overline{0}$ and $\overline{1}$ [4]. Consequently, we need to find approximate solutions.

A subset T of the hypercube is called a binary covering array $\mathrm{CA}(|T|, n, n-k)$ with strength $n-k$ if for each $v \in F_2^n$ and for any k positions there exists $u \in T$

such that v and u can differ only in these k positions. A survey of constructions and bounds for cardinalities of covering arrays can be found in [3]. At this moment, exact bounds are obtained only for small n or for $k = 1, 2, n-2, n-1$ and an arbitrary n. If $n > k+1$, the cardinality of minimum covering array $F(n,k)$ is greater than $2^{n-k-2}(2 + \log(k+2))$ and $F(n,k) \leq 2^n/(k+1)$. Moreover, it is known that $F(n,k) \asymp 2^{n-k} \log n$ as $n \to \infty$ and $n-k$ is fixed. With the exception of the parity check code mentioned above, linear codes are not useful as binary covering arrays. Indeed, the dual code (of cardinality more than two) contains a codeword of weight less $\frac{n}{3}+1$. Then any other proper linear code does not intersect with some $\lceil \frac{n}{3} \rceil$-faces.

We will consider a bit different mathematical problem: to construct a partial covering array $S_k \subset F_2^n$, $|S_k| = 2^{n-k}$ with the following property. The number of k-faces containing elements of S_k is as large as possible. Since in practice we encode elements of set by binary words, it is convenient to use sets of cardinality 2^t. Apparently for the first time partial covering arrays or "covering array with budget constraints"were considered in [2]. In [7] the existence of partial covering arrays with some parameters is established by probabilistic methods. Denote by $\nu_k(S_k)$ the ratio between the number of k-faces that contain elements of S_k and the number $\binom{n}{k}2^{n-k}$ of all k-faces. In Section 2, we prove that $\max \nu_k(S_k) = 1 - \frac{1+o(1)}{2^{k+1}}$ as $n \to \infty$.

For application in information transmission a device or algorithmic function that performs a quantization must be cost-effective. For example, linear codes are easily implemented. In Section 3, we find precise value of $\max \nu_k(S_k)$ for linear sets as $n = 2^k - 1$ and construct corresponding S_k.

Another approach to study suitability of linear codes for similar tasks was developed in [9].

In the beginning, we consider a random subset T of the hypercube with cardinality 2^{n-k}. We suppose that the elements of T are selected independently. Since the probability of any k-face Γ equals $1/2^{n-k}$, the probability $\Pr(T \cap \Gamma = \varnothing)$ equals $(1 - \frac{1}{2^{n-k}})^{2^{n-k}}$. Since $(1 - \frac{1}{2^{n-k}})^{2^{n-k}} \to 1/e$ as $n \to \infty$, the following proposition is true.

Proposition 1. *$\lim_{n\to\infty} \mathrm{E}\nu_k(T) = 1 - 1/e$, where $T \subset E^n$ is a random set, $|T| = 2^{n-k}$, and k is fixed.*

2. Upper bound for partial covering arrays

In this section we will use definitions and methods from [4]. Consider the vector space $\mathbb{V}$ of real-valued functions on F_2^n with the scalar product

$$(f,g) = \frac{1}{2^n} \sum_{x \in F_2^n} f(x)g(x).$$

For every $z \in F_2^n$ define a *character* $\phi_z(x) = (-1)^{\langle x,z \rangle}$, where $\langle x,z\rangle = x_1z_1 + \cdots + x_nz_n$. Here all arithmetic operations are performed on real numbers. As is generally known, the characters form an orthonormal basis of $\mathbb{V}$.

Let M be the adjacency matrix of the hypercube F_2^n. This means that $Mf(x) = \sum_{y:d(x,y)=1} f(y)$, where $d(x,y)$ is the Hamming distance. It is well known that the characters are eigenvectors of M. Indeed, we have

$$M\phi_z(x) = (n - 2wt(z))\phi_z(x),$$

where $wt(z)$ is the number of nonzero coordinates of z.

Let M_r be the adjacency matrix of the distance r, i.e. $M_r(x, y) = 1$ iff $d(x, y) = r$. This matrix generates Bose — Mesner algebra. We have the equations

$$M_r M = (n - r + 1)M_{r-1} + (r + 1)M_{r+1},$$

$$M^r = a_0^r M_0 + a_1^r M_1 + \cdots + a_r^r M_r, \tag{1}$$

where

$$a_i^{r+1} = i a_{i-1}^r + (n - i)a_{i+1}^r, \tag{2}$$

$a_{-1}^r = a_{r+1}^r = 0$. The following properties of the coefficients a_i^r are not difficult to prove by induction using (2).

1. $a_r^r = r!$.
2. $a_i^r = 0$ if $i > r$.
3. $a_i^r = 0$ if r and i have different parity.
4. Consider a_i^r as a function of the variable n. Then $a_i^r(n)$ is a polynomial of degree $(r - i)/2$ as r and i have the same parity. Moreover, $a_i^r(n) = C(r, r - i)n^{(r-i)/2} + p_{r,i}(n)$ where $C(r, s) = \frac{(r+s)!(r-s+1)}{s!(r+1)!}$ are entries of Catalan's triangle and $p_{r,i}$ is some polynomial of the degree at most $\frac{r-i}{2} - 1$.

Notice that a_i^r is the number of the paths of the length r in the hypercube such that distance between the origin and the end of the path is equal to i. It is known (see [4]) that $a_i^r = (\cosh^{n-i}(x)\sinh^i(x))^{(r)}|_{x=0}$ and

$a_i^r = \frac{1}{2^n}\sum\limits_{k=0}^{n}(n - 2k)^r P_k(i, n)$ where

$P_k(r, n) = \sum\limits_{j=0}^{k}(-1)^j\binom{r}{j}\binom{n-r}{k-j}$ are the kth Krawtchouk polynomials in the variable r.

Define $\nu_k(n) = \max \nu_k(S)$ where $S \subset F_2^n$, $|S| = 2^{n-k}$.

Theorem 1. *It holds* $\nu_k(n) \le 1 - \frac{1+o(1)}{2^{k+1}}$ *as* $n \to \infty$.

Proof. Let $r \le k$ be even. For any function $g \in \mathbb{V}$ it holds

$$(M^r g, g) = \sum_{t=0}^{n}(n - 2t)^r \sum_{wt(z)=t}(\phi_z, g)^2 \tag{3}$$

because of $g = \sum\limits_{t=0}^{n}\sum\limits_{wt(z)=t}(\phi_z, g)\phi_z$ and $M\phi_z = (n - 2t)\phi_z$ as $wt(z) = t$.

Consider $S \subset F_2^n$, $|S| = 2^{n-k}$, and let f be the characteristic function of S. Since $(\phi_{\overline{0}}, f) = |S|/2^n = 2^{-k}$, we obtain $(M^r f, f) \ge n^r 2^{-2k}$, i. e. $(M^r f, f)$ is not less than the first term in (3).

It is easy to see that $\sum\limits_x M_i g(x) = \binom{n}{i}\sum\limits_x g(x)$ for any function $g \in \mathbb{V}$. Then it holds $(M_i f, f) \le \binom{n}{i}(f, f)$. Hence $(a_i^r M_i f, f) \le a_i^r\binom{n}{i}/2^k = O(n^{i+(r-i)/2})$ as $n \to \infty$. Since $a_r^r = r!$, from (1) we obtain

$$(M_r f, f) = \left(\frac{1}{a_r^r} + o(1)\right)(M^r f, f) = \frac{n^r(1+o(1))}{r! 2^{2k}}. \tag{4}$$

By definition of M_r the number of the unordered pairs from S at the distance r is equal $\frac{1}{2}\sum_x M_r f(x)\cdot f(x) = 2^{n-1}(M_r f, f)$. Since a pair of vertices at distance r lies in $\binom{n-r}{k-r}$ k-faces Γ, the number of tuples $\langle x, y, \Gamma\rangle$ where $x, y \in S\cap\Gamma$, $d(x,y) = r$, is equal to $2^{n-1}(M_r f, f)\binom{n-r}{k-r}$.

Let $W(x) \subset \Gamma(k)$ be a set of k-faces Γ such that $x \in \Gamma$ and $\Gamma(k)$ is the set of all k-faces. It is easy to see by direct verification that

$$\sum_{x\in S}\chi^{W(x)} \geq \frac{2}{\binom{k}{r}}\sum_{x,y\in S, d(x,y)=r}\chi^{W(x)}\chi^{W(y)},$$

where $\chi^{W(x)} : \Gamma(k) \to \{0,1\}$ is the characteristic function of $W(x)$. By consideration of cases $\sum_{x\in S}\chi^{W(x)}(\gamma) = 0$, $\sum_{x\in S}\chi^{W(x)}(\gamma) = 1$ and $\sum_{x\in S}\chi^{W(x)}(\gamma) \geq 2$ we obtain that

$$\chi^{W(S)} \leq \sum_{x\in S}\chi^{W(x)} - \frac{1}{\binom{k}{r}}\sum_{x,y\in S, d(x,y)=r}\chi^{W(x)}\chi^{W(y)},$$

where $W(S) = \bigcup_{x\in S} W(x)$.

Since the number of tuples $\langle x, y, \Gamma\rangle$ is equal to $\sum_{\gamma\in\Gamma(k)}\sum_{x,y\in S, d(x,y)=r}\chi^{W(x)}(\gamma)\chi^{W(y)}(\gamma)$, we obtain that $|W(S)| \leq 2^{n-k}\binom{n}{k} - \frac{2^{n-1}}{\binom{k}{r}}(M_r f, f)\binom{n-r}{k-r}$.

Then $1 - \nu_k(S) = 1 - \frac{|W(S)|}{|\Gamma(k)|} \geq 1 - \frac{2^{n-1}(M_r f,f)\binom{n-r}{k-r}}{2^{n-k}\binom{n}{k}\binom{k}{r}}$ where $|\Gamma(k)| = 2^{n-k}\binom{n}{k}$ is the number of the k-faces. It is easy to see that $\binom{n-r}{k-r}\binom{n}{r} = \binom{n}{k}\binom{k}{r}$. Thus, the statement follows from (4). □

3. Linear codes

The results of this section were announced in [1].

Proposition 2. *Let C be a linear $(n-k)$-dimensional code over $GF(2)$ and Γ be a k-face. Then*[1] $|C \cap (x \oplus \Gamma)| = 0$ or 2^s *for each $x \in F_2^n$, where s does not depend on x.*

Proof. A system of linear equations over $GF(2)$ has 1 or 0 and 2^s solutions for different right-hand side vectors. □

For a linear code C denote by $\mu_k(C)$ the number of k-faces Γ such that $C \cap \Gamma = \{\overline{0}\}$. By Proposition 2 we obtain that $\nu_k(C) > \mu_k(C)/\binom{n}{k}$.

Let $H = \{h_{ij}\}$ be a binary $k \times n$ matrix. Consider the linear code $C = \{x \in F_2^n | Hx = \overline{0}\}$. The characteristic function χ^C of C is equal to $\prod_i(1 \oplus \bigoplus_j h_{ij}x_j)$. Any Boolean function can be uniquely represented in the algebraic normal form, i. e. as polynomial over $GF(2)$. From known properties of the algebraic normal form we have the following proposition.

Proposition 3. *The number of monomials of the degree k in χ^C is equal to $\mu_k(C)$.*

---

[1] Here and below, $\oplus$ denotes modulo 2 addition.

Let $H(x)$ be the matrix with columns $h_j x_j$ where h_j is a column of H, $j = 1, \dots, n$. For example, if $H_2 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$ then $H_2(x) = \begin{pmatrix} 0 & x_2 & x_3 \\ x_1 & 0 & x_3 \end{pmatrix}$.

The permanent of the matrix H (square or rectangular $k \times n$, $k \le n$) is the sum of all products $h_{1j_1} \dots h_{kj_k}$ where $j_1, \dots, j_k$ are mutually different. For example, $\mathrm{per}_2(H_2(x)) = x_1x_2 \oplus x_2x_3 \oplus x_1x_3$.

By the definition we can obtain the following properties of $\mathrm{per}_2 H(x)$.

1. If we add a row of $H(x)$ to another row of $H(x)$ then $\mathrm{per}_2(H(x))$ does not change.

2. Laplace expansion for determinants is true for the permanents of rectangular matrices (decompositions by row only).

3. If two columns h_{n-1} and h_n of H coincide and $k < n$, then $\mathrm{per}_2(H(x_1, \dots, x_n))$ is equal to $\mathrm{per}_2(H'(x_1, \dots, x_{n-2}, x_{n-1} \oplus x_n))$ where H' is H without the last column.

These properties are proved analogously to the property of the determinant because the permanent and the determinant coincide modulo 2.

We can consider any algebraic normal form as a formal polynomial. Define by $P_H[x_1, \dots, x_n]$ the formal polynomial (where $\cdot$ is a multiplication and $+$ is an addition) corresponding to the algebraic normal form of Boolean function $\mathrm{per}_2(H(x_1, \dots, x_n))$. It is easy to see that $P_H[x_1, \dots, y + z, \dots, x_n]$ corresponds to $\mathrm{per}_2(H(x_1, \dots, y \oplus z, \dots, x_n))$ if z and y do not coincide with any x_i. Let R_H be the real valued function that is obtained from $P_H[x_1, \dots, x_n]$ by substitution real values for variables. From Proposition 3 we obtain

Proposition 4. *Let $H = \{h_{ij}\}$ be a binary $k \times n$ matrix and $C = \{x \in F_2^n | Hx = \overline{0}\}$. Then $\mu_k(C) = R_H(1, \dots, 1)$.*

Let H_k be the matrix $k \times (2^k - 1)$ such that the jth row of H_k is the binary representation $b_k(j)$ of j. We have from the definition

Proposition 5. $P_{H_k}[x_1, \dots, x_n] = \sum\limits_{(i_1, \dots, i_k) \in I} x_{i_1} \cdots x_{i_k}$, *where* $I = \{(i_1, \dots, i_k) \mid i_1 < i_2 < \dots < i_k$ and $b_k(i_1), \dots, b_k(i_k)$ are linearly independent over $GF(2)\}$.

Below we will study the properties of the function

$$f_k(x_1, \dots, x_{2^k-1}) = R_{H_k}(x_1, \dots, x_{2^k-1}).$$

Let H be a binary $k \times n$ matrix having a_j columns $b_k(j)$ for $j = 1 \dots 2^k - 1$. Consider the linear code $C = \{x \in F_2^n | Hx = \overline{0}\}$. Applying properties of permanents we can prove

Proposition 6. $\mu_k(C) = f_k(a_1, \dots, a_{2^k-1})$.

Proof. By the third property of the permanent we have

$$\mathrm{per}_2(H(x_1, \dots, x_n)) = \mathrm{per}_2(H_k(\underbrace{x_{i_1} \oplus \dots}_{a_1}, \dots, \underbrace{x_{i_n} \oplus \dots}_{a_{2^k-1}})).$$

Therefore $P_H[x_1, \dots, x_n] = P_{H_k}[\underbrace{x_{i_1} + \dots}_{a_1}, \dots, \underbrace{x_{i_n} + \dots}_{a_{2^k-1}}]$. Then we obtain equations $\mu_k(C) = R_H(1, \dots, 1) = f_k(a_1, \dots, a_{2^k-1})$ from Proposition 4. $\square$

From Proposition 6 we can conclude that the original problem is equivalent to finding maximum of f_k where the sum of the arguments is fixed. Consider the set

$$T_k(s) = \{(x_1, \dots, x_n) \mid x_j \geq 0, \sum_{j=1}^{n} x_j = s\},$$

where $n = 2^k - 1$.

Theorem 2. $\max\limits_{x \in T_k(s)} f_k(x) = f_k(s/n, \dots, s/n)$.

Proof. Proof by induction on k. For $k = 2$ the statement of the theorem can be verified by the standard methods of analysis. Consider occurrences of the variable x_1 in the polynomial $P_{H_k}[x_1, \dots, x_n]$. By the second and the third properties of the permanent we get

$$\begin{aligned} &\mathrm{per}_2(H_{k+1}(x_1, \dots, x_{2n+1})) \\ &\qquad = x_1 \mathrm{per}_2(H_k(x_{i_1^1} \oplus x_{i_1^2}, \dots, x_{i_1^{2n-1}} \oplus x_{i_1^{2n}})) \oplus g(x_2, \dots, x_{2n+1}) \end{aligned}$$

and $P_{H_{k+1}}[x_1, \dots, x_{2n+1}] = x_1 P_{H_k}[x_{i_1^1} + x_{i_1^2}, \dots, x_{i_1^{2n-1}} + x_{i_1^{2n}}] + G[x_2, \dots, x_{2n+1}]$,

where the polynomial G does not depend on the variable x_1, $b_k(i_1^{2j-1})$ and $b_k(i_1^{2j})$ differ only in the first position. Since each monomial of $P_{H_{k+1}}$ has degree $k+1$ we have

$$P_{H_{k+1}}[x_1, \dots, x_{2n+1}] = \frac{1}{k+1} \sum_{j=1}^{2n+1} x_j P_{H_k}[x_{i_j^1} + x_{i_j^2}, \dots, x_{i_j^{2n-1}} + x_{i_j^{2n}}],$$

where $\{x_{i_j^m}\}_{m=1,\dots,2n}$ is the set of all variables without x_j. Hence we have the equation

$$f_{k+1}(x_1, \dots, x_{2n+1}) = \frac{1}{k+1} \sum_{j=1}^{2n+1} x_j f_k(x_{i_j^1} + x_{i_j^2}, \dots, x_{i_j^{2n-1}} + x_{i_j^{2n}}),$$

where $\{x_{i_j^m}\}_{m=1,\dots,2n}$ is the set of all variables without x_j.

By induction, we find

$$f_{k+1}(x_1, \dots, x_{2n+1}) \leq \frac{1}{k+1} \sum_{j=1}^{2n+1} x_j f_k((s - x_j)/n, \dots (s - x_j)/n).$$

Define $y_j = s - x_j$. Obviously we have $\sum\limits_{j=1}^{2n+1} y_j = 2ns$. Then we obtain the inequality $f_{k+1}(x_1, \dots, x_{2n+1}) \leq \frac{f_k(\bar{1})}{k+1} \sum\limits_{j=1}^{2n+1} (s - y_j) \left(\frac{y_j}{n}\right)^k$.

By the method of Lagrange multipliers, it is not complicated to prove that the function $g(y_1, \dots, y_{2n+1}) = \sum\limits_{j=1}^{2n+1} (s - y_j) y_j^k$ has maximum in the interior point $y_1 = \dots = y_{2n+1} = \frac{2ns}{2n+1}$ if $\sum\limits_{j=1}^{2n+1} y_j = 2ns$, $0 \leq y_j \leq s$. In the edge points (where $y_i = 0$ for some $i \in \{1, \dots, 2n+1\}$), this function isn't positive. Induction step is proved. $\square$

Corollary 1. *For fixed k and $n = m(2^k-1)$ the maximum value of $\mu_k(C_{k,m})$ corresponds to the code $C_{k,m}$ with the check matrix H consisting of m columns $b_k(j)$ for all $j = 1,\dots,2^k-1$. If $m = 1$ then $C_{k,1}$ is the Hamming code.*

For example, $H = (111\dots1)$;

$$H = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix};$$

$$H = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix};$$

$$H = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

By linear algebra it is routine to prove the following proposition.

Proposition 7. *Let $n = 2^k-1$, $k \geq 2$ then $\mu_k(C_{k,1}) = (2^k-1)(2^k-2)(2^k-4)\cdots(2^k-2^{k-1})/k!$ and*

$$\nu_k(C_{k1}) = \frac{1}{k!2^k\binom{n}{k}}\sum_{t=1}^{k-1}(2^k-1)(2^k-2)(2^k-4)\cdots(2^k-2^t)\binom{2^{t+1}-t-2}{k-t-1}2^{t+1}.$$

It is possible to calculate that $\nu_2(C_{2,1}) = 1$, $\nu_3(C_{3,1}) = 9/10$, $\nu_4(C_{4,1}) = 10/13$.

Define $\nu'_k(C) = \mu_k(C)/\binom{n}{k}$. Obviously, $\nu'_k(C)$ is a lower bound of $\nu_k(C)$. Then $\nu'_k(C_{k,1}) = (2^k-1)(2^k-2)(2^k-4)\cdots(2^k-2^{k-1})/(2^k-1)(2^k-2)(2^k-3)\cdots(2^k-k)$ and $\lim\limits_{k\to\infty}\nu'_k(C_{k,1}) \approx 0.2888$.

Corollary 2. $\lim\limits_{s\to\infty}\nu'_k(C_{k,s}) = (2^k-1)(2^k-2)(2^k-4)\cdots(2^k-2^{k-1})/(2^k-1)^k$.

4. Conclusion

As we can see from Proposition 1 the best (for our task) linear code is not better than a random set as k is large. But if k is a small integer than the best linear code is close to the best unrestricted partial covering array. Problems to find the best unrestricted set and to find the asymptotics of its cardinality are open.

The author would like to thank the referee for useful comments.

References

[1] A. Filippov, V. Rufitskiy, V. Potapov, *Scalar-quantization-based multi-layer data hiding for video coding applications*, Visual Communications and Image Processing Conference, 2014, IEEE, 262–265. DOI: 10.1109/VCIP.2014.7051554

[2] A. Hartman, L. Raskin, *Problems and algorithms for covering arrays*, Discrete Math., **284**: 1-3 (2004), 149–156. Zbl 1044.05029

[3] J. Lawrence, R.N. Kacker, Y. Lei, D.R. Kuhn, M. Forbes, *A survey of binary covering arrays*, Electronic J. of Combinatorics, **18**:1 (2011), P84. Zbl 1298.05048

[4] F.J. MacWilliams, N.J.A. Sloane, *The theory of error-correcting codes*, North-Holland, 1977. Zbl 0369.94008

[5] C. Munuera, *Hamming codes for wet paper steganography*, Des. Codes Cryptogr., **76**:1 (2015), 101–111. Zbl 1346.94118

[6] L.H. Ozarow, A.D. Wyner, *Wire-tap channel II*, Advances in cryptology, Lecture Notes in Computer Science, **209** (1985), 33–50. Zbl 0596.94007

[7] K. Sarkar, C. J. Colbourn, A. de Bonis, U. Vaccaro, *Partial Covering Arrays: Algorithms and Asymptotics*, Lecture Notes in Computer Science, **9843** (2016), 437–448. Zbl 06631040

[8] M. Wu, H. Yu, B. Liu, *Data hiding in image and video: part II — design and applications*, IEEE Trans. Image Processing, **12**:6 (2003), 696–705. DOI 10.1109/TIP.2003.810589

[9] Z. Zhuang, B. Dai, Y. Luo, A.J. Han Vinck, *On the relative profiles of a linear code and a subcode*, Des. Codes Cryptogr., **72**:2 (2014), 219–247. Zbl 1333.94063

Vladimir N. Potapov
Sobolev Institute of Mathematics,
pr. Koptyuga, 4,
630090, Novosibirsk, Russia
E-mail address: vpotapov@math.nsc.ru