Des. Codes Cryptogr. (2017) 82:197–209
DOI 10.1007/s10623-016-0265-9

Cover-free codes and separating system codes

A. G. D'yachkov[1] · I. V. Vorobyev[1,2] · N. A. Polyanskii[1,2] · V. Yu. Shchukin[1,2]

Received: 20 September 2015 / Revised: 27 June 2016 / Accepted: 28 July 2016 /
Published online: 19 August 2016

Abstract We give some relations between the asymptotic rates of cover-free (CF) codes, separating system (SS) codes and completely separating system (CSS) codes. We also provide new upper bounds on the asymptotic rate of SS codes based on known results for CF and CSS codes. Finally, we derive a random coding bound for the asymptotic rate of SS codes and give tables of numerical values corresponding to our improved upper bounds.

Keywords Separating system codes · Cover-free codes · Completely separating system codes · Frameproof codes · Digital fingerprinting

Mathematics Subject Classification 94B25 · 94B65

1 Notation, definitions and results

Let q, N, t be integers such that $q \geq 2$, N, the symbol $\triangleq$ denotes equality by definition, $\mathbf{q} \triangleq \{0, 1, \ldots, q-1\}$ – q-ary alphabet, $|A|$ – cardinality of the set A, and $[N] \triangleq \{1, 2, \ldots, N\}$

The material in this work was presented in part at the 2015 IEEE International Symposium on Information Theory [6]. This paper contains some improvements of the results, which were developed in [6] (refer to Theorem 2, Theorem 3 and Statement 3 in Theorem 4).

This is one of several papers published in *Designs, Codes and Cryptography* comprising the "Special Issue on Coding and Cryptography".

✉ A. G. D'yachkov
agd-msu@yandex.ru

✉ I. V. Vorobyev
vorobyev.i.v@yandex.ru

[1] Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, 119992 Moscow, Russian Federation

[2] Institute for Information Transmission Problems, Russian Academy of Sciences, 127051 Moscow, Russian Federation

– the set of integers from 1 to N. The standard symbol $\lfloor a \rfloor$ ($\lceil a \rceil$) will be used to denote the largest (least) integer $\le a$ ($\ge a$). Introduce a q-ary matrix X with t columns (codewords) $\boldsymbol{x}(j)$, $j \in [t]$, and N rows (coordinates) $\boldsymbol{x}_i$, $i \in [N]$, i.e.,

$$X \triangleq \|x_i(j)\|, \ x_i(j) \in \mathbf{q}, \quad \boldsymbol{x}(j) \triangleq (x_1(j), \ldots, x_N(j)), \quad \boldsymbol{x}_i \triangleq (x_i(1), \ldots, x_i(t)). \tag{1}$$

Any such matrix is called a q-ary *code* X of *length* N and *size* t. For $q = 2$, the number of ones in column $\boldsymbol{x}(j)$, i.e., $|\boldsymbol{x}(j)| \triangleq \sum_{i=1}^{N} x_i(j)$, is called the *weight* of $\boldsymbol{x}(j)$, $j \in [t]$. Let Q, $0 < Q < 1$, be a fixed parameter. A binary code X of length N and size t is said to be the *constant-weight* code of *relative* weight Q when $|\boldsymbol{x}(j)| = \lceil Q\, N \rceil$ for any $j \in [t]$.

1.1 q-ary separating system codes

Let $q \ge 2$, $s \ge 1$ and $\ell \ge 1$ be positive integers.

Definition 1 [8] A q-ary code X is called a *q-ary separating system* (SS) (s, ℓ)*-code*, if for any two disjoint sets S, $L \subset [t]$, $|S| \le s$, $|L| \le \ell$, $S \cap L = \varnothing$, there exists an index i, $i \in [N]$, and the corresponding row (coordinate) $\boldsymbol{x}_i$, such that two coordinate sets $\{x_i(j),\ j \in S\} \subseteq \mathbf{q}$ and $\{x_i(j),\ j \in L\} \subseteq \mathbf{q}$ are disjoint. If $\ell = 1$, then q-ary SS $(s, 1)$-codes are also called *frameproof codes*.

The most important applications of SS codes are connected with automata synthesis (see [13,14]), digital fingerprinting (see [1,2]), and constructions of hash functions [17].

Taking into account the obvious symmetry over the parameters s and ℓ, we denote by $t_{ss}^{(q)}(N, s, \ell) = t_{ss}^{(q)}(N, \ell, s)$ the maximal size of SS (s, ℓ)-codes of length N, and by $N_{ss}^{(q)}(t, s, \ell) = N_{ss}^{(q)}(t, \ell, s)$ the minimal length of SS (s, ℓ)-codes of size t. Introduce the *asymptotic rate* of q-ary separating system (s, ℓ)-codes:

$$R_{ss}^{(q)}(s, \ell) = R_{ss}^{(q)}(\ell, s) \triangleq \varlimsup_{N \to \infty} \frac{\log_q t_{ss}^{(q)}(N, s, \ell)}{N}. \tag{2}$$

The purpose of this paper is to obtain new bounds on the asymptotic rate $R_{ss}^{(q)}(s, \ell)$. The case of binary codes is investigated in detail. To compare new bounds of this paper to previous results we provide tables with numerical values in Sect. 1.7.

1.2 Binary separating system codes

Let $s \ge 1$ and $\ell \ge 1$ be integers.

Definition 2 [8] A binary code X is called a *binary separating system (SS)* (s, ℓ)*-code*, if for any two disjoint sets S, $L \subset [t]$, $|S| \le s$, $|L| \le \ell$, $S \cap L = \varnothing$, there exists a row $\boldsymbol{x}_i$, $i \in [N]$, such that

$$x_i(j) = 0 \quad \text{for any} \quad j \in S \quad \text{and} \quad x_i(k) = 1 \quad \text{for any} \quad k \in L$$

or

$$x_i(j) = 1 \quad \text{for any} \quad j \in S \quad \text{and} \quad x_i(k) = 0 \quad \text{for any} \quad k \in L. \tag{3}$$

Definition 2 is equivalent to Definition 1 of q-ary SS codes for $q = 2$.

Definition 3 [10] A binary code X is called a *completely separating system (CSS)* (s, ℓ)*-code*, if for any two disjoint sets S, $L \subset [t]$, $|S| \le s$, $|L| \le \ell$, $S \cap L = \varnothing$, there exist two rows $\boldsymbol{x}_i$, $\boldsymbol{x}_j$, i, $j \in [N]$, such that

$$x_i(m) = 0 \quad \text{for any} \quad m \in S \quad \text{and} \quad x_i(k) = 1 \quad \text{for any} \quad k \in L$$

and

$$x_j(m) = 1 \quad \text{for any} \quad m \in S \quad \text{and} \quad x_j(k) = 0 \quad \text{for any} \quad k \in L.$$

Definition 4 [4,11] A binary code X is called *cover-free (CF)* (s, ℓ)*-code*, if for any two disjoint sets $S,\ L \subset [t]$, $|S| \le s$, $|L| \le \ell$, $S \cap L = \varnothing$, there exists a row $\boldsymbol{x}_i$, $i \in [N]$, such that

$$x_i(j) = 0 \quad \text{for any} \quad j \in S \quad \text{and} \quad x_i(k) = 1 \quad \text{for any} \quad k \in L.$$

Denote by $t_{ss}(N, s, \ell) = t_{ss}(N, \ell, s)$, $t_{css}(N, s, \ell) = t_{css}(N, \ell, s)$, $t_{cf}(N, s, \ell) = t_{cf}(N, \ell, s)$ the maximal size of SS, CSS and CF (s, ℓ)-codes of length N. Also denote by $N_{ss}(t, s, \ell) = N_{ss}(t, \ell, s)$, $N_{css}(t, s, \ell) = N_{css}(t, \ell, s)$, and $N_{cf}(t, s, \ell) = N_{cf}(t, \ell, s)$ the minimal length of the corresponding code of size t. Equalities hold because of the symmetry in parameters s and ℓ. Define the *asymptotic rates* of codes in a standard manner:

$$R_{ss}(s, \ell) = R_{ss}(\ell, s) \triangleq \overline{\lim_{N\to\infty}} \frac{\log_2 t_{ss}(N, s, \ell)}{N} = \overline{\lim_{t\to\infty}} \frac{\log_2 t}{N_{ss}(t, s, \ell)}. \tag{4}$$

$$R_{css}(s, \ell) = R_{css}(\ell, s) \triangleq \overline{\lim_{N\to\infty}} \frac{\log_2 t_{css}(N, s, \ell)}{N} = \overline{\lim_{t\to\infty}} \frac{\log_2 t}{N_{css}(t, s, \ell)}. \tag{5}$$

$$R_{cf}(s, \ell) = R_{cf}(\ell, s) \triangleq \overline{\lim_{N\to\infty}} \frac{\log_2 t_{cf}(N, s, \ell)}{N} = \overline{\lim_{t\to\infty}} \frac{\log_2 t}{N_{cf}(t, s, \ell)}. \tag{6}$$

SS (s, ℓ)-codes and CSS (s, ℓ)-codes (i.e, bounds on the rates (4)-(5) and constructions) have been investigated in many papers. See for instance [3,14].

Let us fix some well-known useful properties.

Proposition 1 [3,14]

1. *For any s and ℓ*

$$R_{ss}(s, l)/2 \le R_{css}(s, l) \le R_{cf}(s, l) \le R_{ss}(s, l). \tag{7}$$

2. *For any s*

$$R_{cf}(s, s) = R_{css}(s, s). \tag{8}$$

Proposition 2 [4]

1. *If we remove from a CF (s, ℓ)-code the column $\boldsymbol{x}(i)$ and all rows j such that $\boldsymbol{x}_j(i) = 1$, then the obtained code is a CF $(s - 1, \ell)$-code.*
2. *If we remove from a CF (s, ℓ)-code the column $\boldsymbol{x}(i)$ and all rows j such that $\boldsymbol{x}_j(i) = 0$, then the obtained code is a CF $(s, \ell - 1)$-code.*

Proposition 3 [4] *The concatenation of a q-ary SS (s, ℓ)-code (as outer code) with a binary CF (s, ℓ)-code (as inner code) is a binary CF (s, ℓ)-code.*

Proposition 4 *The concatenation of a q-ary SS (s, ℓ)-code (as outer code) with a binary SS (s, ℓ)-code (as inner code) is a binary SS (s, ℓ)-code.*

The last proposition was firstly proved in [12] for the particular case of SS (2, 1)-codes. The generalization for the case of arbitrary s and ℓ is straightforward, so we state this result without a proof. Some improvements of inequalities (7) are presented in

Theorem 1 *The rate $R_{ss}(s,\ell)$ of SS (s,ℓ)-codes and the rate $R_{cf}(s,\ell)$ of CF (s,ℓ)-codes satisfy inequalities*

$$\begin{aligned} R_{cf}(s,\ell) \;&\le\; R_{ss}(s,\ell) \le R_{cf}(s-1,\ell), \quad \ell \ge 1,\ s \ge 2,\\ R_{cf}(s,\ell) \;&\le\; R_{ss}(s,\ell) \le R_{cf}(s,\ell-1), \quad \ell \ge 2,\ s \ge 1. \end{aligned} \tag{9}$$

The proof of Theorem 1 as well as the proofs of Theorems 2–4 is given in Sect. 2. The best known upper bounds on the rate $R_{cf}(s,\ell)$ of CF (s,ℓ)-codes were established in [9]. These bounds and inequalities from Theorem 1 lead to new upper bounds on the rate $R_{ss}(s,\ell)$ of binary SS (s,ℓ)-codes which improve the previously known upper bounds for many specific values of parameters s and ℓ (see Table 2).

1.3 Asymptotic bounds on the rates of CF (s,ℓ)-codes and binary SS (s,ℓ)-codes

For fixed $\ell \ge 2$ and $s \to \infty$ the best known asymptotic upper bound on the rate $R_{cf}(s,\ell)$ was proved in [5]. This bound has the following form:

$$R_{cf}(s,\ell) \;\leqslant\; \frac{(\ell+1)^{\ell+1}}{2e^{\ell-1}}\,\frac{\log_2 s}{s^{\ell+1}}\,(1+o(1)). \tag{10}$$

From Theorem 1 it follows that the rate $R_{ss}(s,\ell)$ of binary SS (s,ℓ)-codes satisfies the same asymptotic inequality (10).

In paper [5] we published an asymptotic lower bound on the rate $R_{cf}(s,\ell)$ which has the same asymptotic behavior as the right-hand side of Upper bound (10). Unfortunately, our proof in [5] based on considerations of the ensemble of binary constant-weight codes is incorrect. At present the best known asymptotic lower bound established in [4] is

$$R_{cf}(s,\ell) \ge \frac{\ell^{\ell}}{e^{\ell}}\,\frac{\log_2 e}{s^{\ell+1}}\,(1+o(1)). \tag{11}$$

1.4 q-ary separating system codes

Formulations of our new results about non-asymptotic connections of the rate (2) of q-ary SS (s,ℓ)-codes with the rates (4) and (6) of binary codes are summarized in

Theorem 2 *Let $q \ge 2$, $m = \min(\max(q-s,1),\ell)$, $n = \min(\max(q-\ell,1),s)$. Then the rate $R_{ss}^{(q)}(s,\ell)$ of q-ary SS codes satisfies inequalities:*

$$R_{ss}^{(q)}(s,\ell) \;\le\; \frac{(2^{q-1}-1)\cdot R_{ss}(s,\ell)}{\log_2 q}, \tag{12}$$

$$R_{ss}^{(q)}(s,\ell) \;\le\; \frac{\sum_{k=m}^{\ell}\binom{q-1}{k}\cdot R_{cf}(s-1,\ell)}{\log_2 q} \quad \textit{for } s \geqslant 2, \tag{13}$$

$$R_{ss}^{(q)}(s,\ell) \;\le\; \frac{\sum_{k=m}^{\ell}\binom{q-1}{k-1}\cdot R_{cf}(s,\ell-1)}{\log_2 q} \quad \textit{for } \ell \geqslant 2, \tag{14}$$

$$R_{ss}^{(q)}(s,\ell) \;\le\; \frac{\sum_{k=n}^{s}\binom{q-1}{k}\cdot R_{cf}(s,\ell-1)}{\log_2 q} \quad \textit{for } \ell \geqslant 2, \tag{15}$$

$$R_{ss}^{(q)}(s,\ell) \;\le\; \frac{\sum_{k=n}^{s}\binom{q-1}{k-1}\cdot R_{cf}(s-1,\ell)}{\log_2 q} \quad \textit{for } s \geqslant 2. \tag{16}$$

The following lower bound on the rate $R_{ss}^{(q)}(s,\ell)$ was proved in paper [17]

$$R_{ss}^{(q)}(s,\ell) \ \geq \ \frac{\lfloor \log_2 q \rfloor \, R_{ss}(s,\ell)}{\log_2 q}. \tag{17}$$

For the particular case $q = 3$, the best numerical values of upper bounds (12)–(16) are given in Table 3.

1.5 Lower bound on the rate of q-ary separating system codes

Theorem 3 (Random coding bound) *For any fixed q, $2 \leq q \leq s+\ell+1$, the rate $R_{ss}^{(q)}(s,\ell)$ of q-ary SS (s,ℓ)-codes satisfies the inequality*

$$R_{ss}^{(q)}(s,\ell) \geq \frac{-\log_q \left(1-(q-1)^\ell \frac{s^s}{(s+\ell)^{s+\ell}}\right)}{s+\ell}. \tag{18}$$

As an asymptotic consequence, for any fixed $q \geq 2$, $\ell \geq 1$ and $s \to \infty$ the rate satisfies

$$R_{ss}^{(q)}(s,\ell) \geq \frac{(q-1)^\ell}{e^\ell \ln q} \frac{1}{s^{\ell+1}}(1+o(1)). \tag{19}$$

The proof of Theorem 3 is based on an ensemble of q-ary codes with independent and identically distributed components which earlier was suggested in [15] to obtain a random coding lower bound on the rate $R_{ss}^{(q)}(s,1)$ for the particular case of q-ary SS $(s,1)$-codes called frameproof codes. For small s, ℓ and q the non-asymptotic lower bound (18) could be essentially improved by a proper choice of the ensemble parameters.

If $q \geq 2$ and $\ell \geq 1$ are fixed, then combining (10) and (13), one can easily obtain the following asymptotic ($s \to \infty$) upper bound on the rate $R_{ss}^{(q)}(s,\ell)$ of q-ary SS codes:

$$R_{ss}^{(q)}(s,\ell) \ \leq \ \frac{\sum_{k=1}^{\ell} \binom{q-1}{k}(\ell+1)^{\ell+1}}{2e^{\ell-1}\log_2 q} \cdot \frac{\log_2 s}{s^{\ell+1}}\,(1+o(1)), \quad s \to \infty. \tag{20}$$

The ratio of the upper bound (20) to the lower bound (19) is

$$\frac{\sum_{k=1}^{\ell} \binom{q-1}{k}(\ell+1)^{\ell+1}}{2e^{-1}(q-1)^\ell} \ln s(1+o(1)) \leq \frac{e(\ell+1)^{\ell+1}}{2(\ell-1)!} \ln s(1+o(1)), \quad s \to \infty, \tag{21}$$

Where the inequality in (21) takes place for sufficiently large q, $q > 2\ell$. In the particular case $\ell = 1$, i.e., for q-ary frameproof codes, the right-hand side of (21) equals $2e \ln s(1+o(1))$ and does not depend on q, $q > 2$.

1.6 Recurrent inequalities

The best known upper bounds on the rate $R_{cf}(s,\ell)$ of CF (s,ℓ)-codes are based on the recurrent inequality [7]:

$$R_{cf}(s,\ell) \leq R_{cf}(s-u,\ell-v) \cdot \frac{u^u v^v}{(u+v)^{u+v}}, \quad 1 \leq u \leq s-1, \quad 1 \leq v \leq \ell-1. \tag{22}$$

and its improvement [9]:

$$R_{cf}(s,\ell) \leq \frac{R_{cf}(s-u,\ell-v)}{R_{cf}(s-u,\ell-v) + \frac{(u+v)^{u+v}}{u^u v^v}}, \quad 1 \leq u \leq s-1, \quad 1 \leq v \leq \ell-1. \tag{23}$$

Table 1 Upper bounds on the rate $R_{css}(s, \ell)$ for CSS (s, ℓ)-codes

$s \mid \ell$	1	2	3	4	5
1	1	0.322	0.199	0.14	0.106
2	0.322	0.161	0.0662	0.0429	0.0286
3	0.199	0.0662	0.0353	0.0153	0.0101
4	0.14	0.0429	0.0153	0.00836	0.00370
5	0.106	0.0286	0.0101	0.00370	0.00204
6	0.083	0.0203	0.00669	0.00245	0.000911

The similar joint recurrent inequalities for the rates $R_{cf}(s, \ell)$, $R_{ss}(s, \ell)$ and $R_{css}(s, \ell)$ are formulated below in the form of Theorem 4.

Theorem 4 (Recurrent inequalities)

(1) *For any* $u \in [s-1]$, $v \in [\ell - 1]$,

$$R_{ss}(s, \ell) \le R_{ss}(s-u, \ell - v) \cdot \max_{0 \le z \le 1} \{z^u (1-z)^v + (1-z)^u z^v\}. \tag{24}$$

(2) *For any* $v \in [\ell - 1]$ *and* $u = v + s - \ell$, $s - (\ell - 1) \le u \le s - 1$,

$$R_{ss}(s, \ell) \le R_{css}(s-u, \ell - v) \cdot \max_{0 \le z \le 1} \{z^u (1-z)^v + (1-z)^u z^v\}. \tag{25}$$

(3) *For any* $v \in [\min(s, \ell) - 1]$,

$$R_{ss}(s, \ell)/2 \le R_{css}(s, \ell) \le R_{css}(s-v, \ell - v) \frac{1}{2^{2v}}. \tag{26}$$

Statements (1) and (2) allow us to improve some upper bounds on the rate $R_{ss}(s, \ell)$. Numerical values for small parameters s and ℓ are provided in Table 2. Indeed, if $u = v$, then Statement (3) is stronger than Statements (1) and (2), but for small values of s and ℓ, it does not give any improvement for $R_{css}(s, \ell)$ and $R_{ss}(s, \ell)$. Theorem 4 allows to conclude the following asymptotic behavior:

$$R_{ss}(s, s) \le \frac{\text{const}}{2^{2s}}, \quad s \to \infty.$$

1.7 Tables of upper bounds

In Table 1 we present the best known upper bounds [3] on the rate of CSS (s, ℓ)-codes. Using these values, we put in Table 2 upper bounds on the rate of SS (s, ℓ)-codes calculated with the help of inequalities from Theorem 4.

In Table 2, the *improved* upper bounds on the rate of SS (s, ℓ)-codes are marked by the boldface type.

Let us demonstrate how these values have been obtained. Consider for instance upper bounds for SS $(4, 6)$-codes. Using the second inequality from Theorem 4 with $v = 3$ and $u = 1$ we obtain the following

$$R_{ss}(4, 6) \le R_{css}(3, 3) \max_{0 \le z \le 1} \{z(1-z)^3 + (1-z)z^3\}.$$

The maximum value $\frac{1}{8}$ of $z(1-z)^3 + (1-z)z^3$ is attained at $z = \frac{1}{2}$. Hence, the rate

$$R_{ss}(4, 6) \le \frac{R_{css}(3, 3)}{8} \le \frac{0.0353515}{8} \approx 0.00442.$$

Table 2 Upper bounds on the rate $R_{ss}(s, \ell)$ for SS (s, ℓ)-codes

$s \mid \ell$	1	2	3	4	5
1	1	0.5[c]	**0.322**[a]	**0.199**[a]	**0.14**[a]
2	0.5[c]	0.2835[d]	0.1202[c]	**0.0744**[a]	**0.0455**[a]
3	**0.322**[a]	0.1202[c]	0.06627[c]	0.02951[c]	**0.0183**[a]
4	**0.199**[a]	**0.0744**[a]	0.02951[c]	0.01630[c]	0.00728[c]
5	**0.14**[a]	**0.0455**[a]	**0.0183**[a]	0.00728[c]	0.004037[c]
6	**0.106**[a]	**0.0286**[a]	**0.0109**[a]	**0.00442**[b]	0.00181[c]

[a] Inequality (9)
[b] Statement 2 of Theorem 4
[c] See [3]
[d] See [14]

Table 3 Upper bounds on the rate $R_{ss}(s, \ell)$ for ternary SS (s, ℓ)-codes

(s, ℓ)	Old [3]	New	(s, ℓ)	Old [3]	New
(2, 2)	**0.3537**	0.5366	(4, 3)	0.07056	**0.05586**
(3, 3)	**0.1138**	0.1254	(5, 4)	0.02290	**0.01378**
(4, 4)	0.03675	**0.0308**	(4, 2)	0.1605	**0.1408**
(5, 5)	0.01202	**0.00764**	(5, 3)	0.05167	**0.03464**
(3, 2)	**0.2197**	0.2275	(5, 2)	0.1268	**0.08612**

Better bounds are in bold

In Table 3 we present upper bounds on the rate of ternary SS (s, ℓ)-codes. For comparison the best previously known values are given.

2 Proofs of Theorems

Proof (Theorem 1) The left-hand side of (9) arises from (7). To prove the right-hand side of (9), we consider an arbitrary SS (s, ℓ)-code X of size t and length N. Construct the code $X' = (\boldsymbol{x}'(1), \boldsymbol{x}'(2), \ldots, \boldsymbol{x}'(t))$ of size t and length $2N$ as follows: $\boldsymbol{x}'(i) \triangleq \boldsymbol{x}(i)^\frown\overline{\boldsymbol{x}(i)}$, $i \in [t]$, where $^\frown$ denotes the concatenation of two vectors, and $\overline{\boldsymbol{x}(i)} \triangleq (\overline{x_1(i)}, \overline{x_2(i)}, \ldots, \overline{x_N(i)})$ denotes the opposite vector to $\boldsymbol{x}(i)$, i.e.,

$$\overline{x_j(i)} \triangleq \begin{cases} 1 & \text{if } x_j(i) = 0, \\ 0 & \text{if } x_j(i) = 1, \end{cases} \quad i \in [t], \quad j \in [N].$$

One can easily verify that X' is a constant weight CF (s, ℓ)-code of relative weight $1/2$. Construct a new code X'' from X' by removing a column of X' and all rows having 1's in this column. From the first statement of Proposition 2 it follows that the obtained code X'' is a CF $(s-1, \ell)$-code of size $t-1$ and length N. Thus $t_{cf}(N, s-1, \ell) \geq t_{ss}(N, s, \ell) - 1$ and $R_{cf}(s-1, \ell) \geq R_{ss}(s, \ell)$. The right-hand side of the first inequality (9) is proved. The similar arguments using the second statement of Proposition 2 yield the right-hand side of the second inequality (9). □

Proof (Theorem 2) To establish Inequalities (12)–(16) we will use a concatenated construction. Consider an arbitrary q-ary SS (s, ℓ)-code X' of size t and length N. This code will be an outer code in our concatenated construction. We use two different inner codes.

1. Construct a $(2^{q-1} - 1) \times q$ matrix, the rows of which are all nonzero binary sequences with zero at the first position. Columns of this matrix form a code D_1 of size q and length

$2^{q-1}-1$. It is easy to see that the code D_1 is a SS (s,ℓ)-code for every s and ℓ. The concatenated code X'' with the inner code D_1 and the outer code X' is a binary SS (s,ℓ)-code (Proposition 4). The rate of the code X'' is equal to

$$\frac{\log_q t}{N}\,\frac{\log_2 q}{2^{q-1}-1},$$

hence Inequality (12) holds.

2. We prove only Inequalities (13) and (14). Inequalities (15) and (16) follow from the symmetry on parameters s and ℓ. Let us denote $C(q,\ell)=\sum_{k=m}^{\ell}\binom{q}{k}$. Consider a $C(q,\ell)\times q$ matrix, which rows are all nonzero binary sequences with p ones, $m\le p\le \ell$ ones. Columns of this matrix form a code D_2 of size q and length $C(q,\ell)$. It is easy to see that the code D_2 is a binary CF (s,ℓ)-code. Indeed, fix two arbitrary disjoint sets $S\subset[t]$, $|S|\le s$, and $L\subset[t]$, $|L|\le\ell$. Consider set $L'\subset[t]$, $m\le|L'|\le\ell$, such that $L\subset L'$ and $L'\cap S=\varnothing$. By construction of the code D_2 there exists a row $\boldsymbol{x}_i$ such that

$$x_i(j)=1 \quad \text{for any} \quad j\in L' \quad \text{and} \quad x_i(k)=0 \quad \text{for any} \quad k\in[t]\setminus L'.$$

Therefore, the code D_2 is a binary CF (s,ℓ)-code. The concatenated code X'' with the inner code D_2 and the outer code X' is a binary CF (s,ℓ)-code (Proposition 3). Furthermore, X'' is a constant weight code with relative weight

$$Q=\frac{\sum_{k=m}^{\ell}\binom{q}{k}\frac{k}{q}}{\sum_{k=m}^{\ell}\binom{q}{k}}=\frac{\sum_{k=m}^{\ell}\binom{q-1}{k-1}}{\sum_{k=m}^{\ell}\binom{q}{k}}.$$

Using Proposition 2 we obtain (13) and (14). □

Proof [Theorem 3] Given a q-ary code X we say that a set U, $U\subset[t]$, having the size $|U|=s+\ell$, is an (s,ℓ)-*bad* set for the code X if there exists a set S, $S\subset U$, of size $|S|=s$, and the corresponding set $L\triangleq U\setminus S$ of size $|L|=\ell$, such that the code X does not contain a row $\boldsymbol{x}_i$, $i\in[N]$, in which the coordinate sets $\{x_i(j),\ j\in S\}\subseteq\mathbf{q}$ and $\{x_i(j),\ j\in L\}\subseteq\mathbf{q}$ are disjoint. □

Given a parameter p, $0<p<1/(q-1)$, consider the following ensemble of $(N\times t)$-matrixes $X=\|x_i(j)\|$, where each q-ary entry $x_i(j)$, $i\in[N]$, $j\in[t]$, is chosen independently and the probability

$$\Pr\{x_i(j)=k\}\triangleq p,\quad k=0,1,\dots q-2,\qquad \Pr\{x_i(j)=q-1\}\triangleq 1-p(q-1). \quad (27)$$

Below we will apply the conventional random coding arguments as the evident

Proposition 5 *If for the ensemble* (27)*, the mathematical expectation of the number of all* (s,ℓ)*-bad sets* <1*, then there exists a* q*-ary code* X*, for which there is no* (s,ℓ)*-bad set, i.e., there exists a SS* (s,ℓ)*-code of size* t *and length* N*.*

Obviously, in the ensemble (27), the probability $P_0(s,\ell)$ of the event "a fixed set U is (s,ℓ)-bad for code X" depends only on parameters (s,ℓ) and does not depend on the set U. We want to construct an upper bound on $P_0(s,\ell)$ as follows. For brevity, denote by $a_1,\dots a_s$ and $b_1,\dots b_l$ the element of a fixed row of X having indexes which belong to the sets S and L respectively. If $P_1(s,\ell)$ is the probability of the event "$a_i\ne b_j$ for $1\le i\le s$, $1\le j\le\ell$", then the inequality

$$P_0(s,\ell)\le\binom{s+\ell}{s}\left(1-P_1(s,\ell)\right)^N.$$

holds. In addition, the conditional probability of the event "$a_i \neq b_j$ for $1 \leq i \leq s, 1 \leq j \leq \ell$" provided that "$b_i \neq q-1$" is not less than $(1-p\ell)^s$. Therefore,

$$P_1(s,\ell) \geq (q-1)^\ell p^\ell (1-p\ell)^s$$

and

$$P_0(s,\ell) \leq \binom{s+\ell}{s}\left(1-(q-1)^\ell p^\ell (1-p\ell)^s\right)^N. \tag{28}$$

The mathematical expectation of the number of all (s,ℓ)-bad sets does not exceed $t^{s+\ell} P_0(s,\ell)$. Hence, Proposition 5 implies that the inequality

$$t^{s+\ell}\binom{s+\ell}{s}\left(1-(q-1)^\ell p^\ell (1-p\ell)^s\right)^N \leq 1. \tag{29}$$

is a sufficient condition for the existence of SS (s,ℓ)-codes. The rate definition (2) and the sufficient condition (29) lead to the lower bound

$$R_{ss}^{(q)}(s,\ell) \geq \frac{-\log_q\left(1-(q-1)^\ell p^\ell (1-p\ell)^s\right)}{s+\ell}.$$

which is true for any value of the parameter p, $0 < p < 1/(q-1)$. If $q \leq s+\ell+1$, then the maximum of $p^\ell(1-p\ell)^s$ is attained at $p = \frac{1}{s+\ell}$. Therefore, the non-asymptotic lower bound (18) on the rate $R_{ss}^{(q)}(s,\ell)$ of SS (s,ℓ)-codes is proved.

For fixed q, ℓ and $s \to \infty$

$$\frac{-\log_q\left(1-(q-1)^\ell \frac{s^s}{(s+l)^{s+l}}\right)}{s+\ell} = \frac{(q-1)^\ell}{s^{\ell+1}\ln q}\frac{s^s}{(s+l)^s}(1+o(1)) = \frac{(q-1)^\ell}{e^\ell \ln q}\frac{1}{s^{\ell+1}}(1+o(1)).$$

Thus, the asymptotic lower bound (19) on the rate $R_{ss}^{(q)}(s,\ell)$ is established.

Proof (Theorem 4) For any integers t and u, $1 \leq u < t$, introduce the set

$$\mathscr{P}_u(t) \triangleq \{P : P \subset [t], \quad |P| = u\}.$$

i.e., the symbol $\mathscr{P}_u(t)$ denotes the collection of all u-subsets of the set $[t]$. Obviously, its size is $|\mathscr{P}_u(t)| = \binom{t}{u}$. Let U, $U \in \mathscr{P}_u(t)$, and V, $V \in \mathscr{P}_v(t)$, where $U \cap V = \varnothing$, be two arbitrary disjoint subsets of the set $[t]$ with cardinalities u and v respectively. For a fixed binary code X of size t and length N, introduce the set of indexes $D_{u,v}(U,V,X)$, $D_{u,v}(U,V,X) \subset [N]$, $0 \leq |D_{u,v}(U,V,X)| \leq N$, such that an index $i \in D_{u,v}(U,V,X)$ if and only if the corresponding row $\boldsymbol{x}_i = (x_i(1), x_i(2), \ldots, x_i(t))$, $i \in [N]$, of the code X satisfies one of the following two conditions:

$$x_i(j) = 0 \text{ for any } j \in U \quad \text{and} \quad x_i(k) = 1 \text{ for any } k \in V$$

or

$$x_i(j) = 1 \text{ for any } j \in U \quad \text{and} \quad x_i(k) = 0 \text{ for any } k \in V. \tag{30}$$

Define the average number

$$\overline{D}_{u,v}(X) \triangleq \sum_{\substack{U \in \mathscr{P}_u(t),\, V \in \mathscr{P}_v(t),\\ U \cap V = \varnothing}} \frac{|D_{u,v}(U,V,X)|}{\binom{t}{u+v}\cdot\binom{u+v}{u}}, \tag{31}$$

and

$$\overline{D}_{u,v}(t, N) = \max_{X} \overline{D}_{u,v}(X),$$

where the maximum is taken over all binary codes X of size t and length N. □

Lemma 1 *The number $\overline{D}_{u,v}(t, N)$ satisfies the asymptotic inequality*

$$\overline{\lim_{t\to\infty}} \frac{\overline{D}_{u,v}(t, N(t))}{N(t)} \le \max_{0\le z\le 1} \{z^u(1-z)^v + (1-z)^u z^v\}, \tag{32}$$

where $N(t)$ is an arbitrary function.

Proof (Lemma 1) Given a subset $K \subset [t]$, $|K| = u + v$, and an index $i \in N$ denote by $\boldsymbol{x}_i(K)$ the $[1 \times (u + v)]$-submatrix of X composed of elements of the i-th row and columns having indexes from the set K. If

$$I^i_{u,v}(K, X) \triangleq \begin{cases} 2 & \text{if } u = v \text{ and } \boldsymbol{x}_i(K) \text{ contains } u = v \text{ zeroes,} \\ 1 & \text{if } u \ne v \text{ and } \boldsymbol{x}_i(K) \text{ contains either } u \text{ zeroes or } v \text{ zeroes,} \\ 0 & \text{otherwise,} \end{cases} \tag{33}$$

then the sum

$$M_{u,v}(X) \triangleq \sum_{i\in N,\, K\in\mathscr{P}_{u+v}(t)} I^i_{u,v}(K, X) \tag{34}$$

can be interpreted as the number of all possible $[1 \times (u + v)]$-submatrices of X, which contain u zeroes and v ones or v zeroes and u ones (submatrices with $u = v$ zeroes are counted twice). If a_i $(t - a_i)$ denotes the number of zeroes (ones) in the ith row of the code X, then from (33)–(34) it follows that the number

$$M_{u,v}(X) = \sum_{i=1}^{N} \binom{a_i}{u} \cdot \binom{t-a_i}{v} + \sum_{i=1}^{N} \binom{a_i}{v} \cdot \binom{t-a_i}{u}.$$

On the other hand, definition (31) implies that

$$M_{u,v}(X) = \overline{D}_{u,v}(X) \cdot \binom{t}{u+v}\binom{u+v}{u}.$$

The last two equations lead to the inequality:

$$\binom{t}{u+v}\binom{u+v}{u} \cdot \overline{D}_{u,v}(X) \le N \cdot \max_{a\in[t]} \left\{ \binom{a}{u} \cdot \binom{t-a}{v} + \binom{a}{v} \cdot \binom{t-a}{u} \right\},$$

which can be written in the form:

$$\frac{\overline{D}_{u,v}(X)}{N} \le \frac{\max_{a\in[t]} \{a^u(t-a)^v + a^v(t-a)^u\}}{(t-(u+v))^{u+v}} = \frac{\max_{a\in[t]} \left\{ \left(\frac{a}{t}\right)^u \left(1-\frac{a}{t}\right)^v + \left(\frac{a}{t}\right)^v \left(1-\frac{a}{t}\right)^u \right\}}{\left(1-\frac{u+v}{t}\right)^{u+v}}.$$

If $t \to \infty$, then the evident passage to the limit yields the asymptotic inequality (32). □

To complete the proof of Statement 1 of Theorem 4, we need

Lemma 2 *For any $u \in [s-1]$ and $v \in [\ell-1]$, the minimal length of SS $(s-u, \ell-v)$-codes of size $t - (u + v)$ satisfies the inequality*

$$N_{ss}(t - (u + v), s - u, \ell - v) \le \overline{D}_{u,v}(t, N_{ss}(t, s, \ell)). \tag{35}$$

Proof (Lemma 2) Consider an arbitrary SS (s, ℓ)-code X of size t and length $N = N_{ss}(t, s, \ell)$. In virtue of definition (31), there exist two disjoint sets $U \subset [t]$, $|U| = u$, and $V \subset [t]$, $|V| = v$, $U \cap V = \emptyset$, such that

$$|D_{u,v}(U, V, X)| \le \overline{D}_{u,v}(X) \le \overline{D}_{u,v}(t, N_{ss}(t, s, \ell)). \tag{36}$$

Define the code X' of length $|D_{u,v}(U, V, X)|$ and size $t-(u+v)$ as a subcode of X containing all rows $\boldsymbol{x}_i$, $i \in D_{u,v}(U, V, X)$ and all columns $\boldsymbol{x}(j)$, $j \in [t] \setminus \{U \cup V\}$. Let us show that X' is a SS $(s-u, \ell-v)$-code. Indeed, fix two arbitrary disjoint sets

$$U', V' \subset [t] \setminus \{U \cup V\}, \quad |U'| = s-u, \quad |V'| = \ell - v, \quad U' \cap V' = \emptyset,$$

and consider two disjoint sets $\{U' \cup U\}, \{V' \cup V\} \subset [t]$.

Obviously, the size $|U' \cup U| = s$ and the size $|V \cup V'| = \ell$.

Hence, the condition (3) from Definition 2 and definition (30) imply that there exists an index i, $i \in D_{u,v}(U, V, X)$, such that either

$$x_i(j) = 0 \quad \text{for any } j \in U' \cup U \quad \text{and} \quad x_i(k) = 1 \quad \text{for any} \quad k \in V' \cup V$$

or

$$x_i(j) = 1 \quad \text{for any } j \in U' \cup U \quad \text{and} \quad x_i(k) = 0 \quad \text{for any} \quad k \in V' \cup V$$

holds. Therefore, the code X' of length $|D_{u,v}(U, V, X)|$, satisfying the inequality (36), is an SS $(s-u, \ell-v)$-code. This conclusion completes the proof of Lemma 2. □

The Inequality (35) of Lemma 2 is equivalent to the inequality

$$\frac{N_{ss}(t-(u+v), s-u, \ell-v)}{N_{ss}(t, s, \ell)} \le \frac{\overline{D}_{u,v}(t, N_{ss}(t, s, \ell))}{N_{ss}(t, s, \ell)}. \tag{37}$$

Letting t tend to infinity and applying Inequality (32) we obtain

$$\begin{aligned}\frac{R_{ss}(s, \ell)}{R_{ss}(s-u, \ell-v)} &= \varlimsup_{t\to\infty} \frac{N_{ss}(t-(u+v), s-u, \ell-v)}{N_{ss}(t, s, \ell)} \\ &\le \varlimsup_{t\to\infty} \frac{\overline{D}_{u,v}(t, N_{ss}(t, s, \ell))}{N_{ss}(t, s, \ell)} \le \max_{0\le z\le 1} \{z^u(1-z)^v + (1-z)^u z^v\}.\end{aligned}$$

Statement 1 of Theorem 4 is proved.

Proof (Statement 2) Proof of the Statement 2 is essentially the same as in Statement 1, but instead of Lemma 2 we need □

Lemma 3 *For any $v \in [\ell - 1]$ and $u = v + s - \ell$, $s - (\ell - 1) \le u \le s - 1$, the minimal length of CSS $(s-u, \ell-v)$-codes of size $t-(u+v)$ satisfies the inequality*

$$N_{css}(t-(u+v), s-u, \ell-v) \le \overline{D}_{u,v}(t, N_{ss}(t, s, \ell)).$$

Proof (Lemma 3) Repeating the arguments from the proof of Lemma 2, consider an arbitrary SS (s, ℓ)-code X of size t and length $N = N_{ss}(t, s, \ell)$, along with two disjoint sets $U \subset [t]$, $|U| = u$, and $V \subset [t]$, $|V| = v$, $U \cap V = \emptyset$, such that the inequality

$$|D_{u,v}(U, V, X)| \le \overline{D}_{u,v}(X) \le \overline{D}_{u,v}(t, N_{ss}(t, s, \ell))$$

holds. Without loss of generality, we assume that for any index i, $i \in D_{u,v}(U, V, X)$,

$$x_i(k) = 0 \quad \text{for any } k \in U \quad \text{and} \quad x_i(k) = 1 \quad \text{for any } k \in V. \tag{38}$$

Define the code X' of length $|D_{u,v}(U, V, X)|$ and size $t - (u + v)$ as a subcode of X containing all rows $\boldsymbol{x}_i$, $i \in D_{u,v}(U, V, X)$ and all columns $\boldsymbol{x}(j)$, $j \in [t] \setminus \{U \cup V\}$. Let us prove that X' is a CSS $(s - u, \ell - v)$-code.

Indeed, fix two arbitrary disjoint sets

$$U', V' \subset [t] \setminus \{U \cup V\}, \quad |U'| = s - u, \quad |V'| = \ell - v, \quad U' \cap V' = \emptyset,$$

and consider two disjoint sets $\{U' \cup U\}, \{V' \cup V\} \subset [t]$, $|U' \cup U| = s$, $|V \cup V'| = \ell$. The condition (3) from Definition 2 and definition (30) imply that there exists an index i, $i \in D_{u,v}(U, V, X)$, such that either

$$x_i(k) = 0 \quad \text{for any } k \in U' \cup U \quad \text{and} \quad x_i(m) = 1 \quad \text{for any} \quad m \in V' \cup V$$

or

$$x_i(k) = 1 \quad \text{for any } k \in U' \cup U \quad \text{and} \quad x_i(m) = 0 \quad \text{for any} \quad m \in V' \cup V$$

holds. From (38) it follows that

$$x_i(k) = 0 \quad \text{for any } k \in U' \cup U \quad \text{and} \quad x_i(m) = 1 \quad \text{for any} \quad m \in V' \cup V,$$

therefore,

$$x_i(k) = 0 \quad \text{for any } k \in U' \quad \text{and} \quad x_i(m) = 1 \quad \text{for any} \quad m \in V'. \tag{39}$$

Considering two disjoint sets $\{U' \cup V\}, \{V' \cup U\} \subset [t]$ of sizes $|U' \cup V| = \ell$, $|U \cup V'| = s$, we obtain that there exists an index j, $j \in D_{u,v}(U, V, X)$, such that either

$$x_j(k) = 1 \quad \text{for any } k \in U' \cup V \quad \text{and} \quad x_j(m) = 0 \quad \text{for any} \quad m \in V' \cup U$$

or

$$x_j(k) = 0 \quad \text{for any } k \in U' \cup V \quad \text{and} \quad x_j(m) = 1 \quad \text{for any} \quad m \in V' \cup U.$$

Condition (38) leads to

$$x_j(k) = 1 \quad \text{for any } k \in U' \cup V \quad \text{and} \quad x_j(m) = 0 \quad \text{for any} \quad m \in V' \cup U,$$

thus,

$$x_j(k) = 1 \quad \text{for any } k \in U' \quad \text{and} \quad x_j(m) = 0 \quad \text{for any} \quad m \in V'. \tag{40}$$

From (39) and (40) it follows that the code X' is a CSS $(s - u, \ell - v)$-code. This completes the proof of Lemma 3 and Statement 2. □

Proof (Statement 3) To proof Statement 3 we need Lemma 4. □

Lemma 4 *For any $v \in [\ell - 1]$ the minimal length of every CSS $(s - v, \ell - v)$-code of size $t - 2v$ satisfies the inequality*

$$2N_{css}(t - 2v, s - v, \ell - v) \le \overline{D}_{v,v}(t, N_{css}(t, s, \ell)).$$

We omit the proof of this Lemma since it is very similar to the proof of Lemma 3.

Theorem 4 is proved.

3 Conclusions

In this paper, we study SS (s, ℓ)-codes. New recurrent inequalities on the asymptotic rates of binary SS (s, ℓ)-codes are given. To prove these inequalities we derive a Plotkin-type bound for SS (s, ℓ)-codes. Also we establish a connection between the asymptotic rate of q-ary SS (s, ℓ)-codes and the asymptotic rates of binary SS (s, ℓ)-codes, which allows us to improve upper bounds for q-ary SS (s, ℓ)-codes. Tables of numerical values for the improved upper bounds are presented for$q = 2; 3$. A random coding bound on the rate of q-ary SS (s, ℓ)-codes is proved. We check that for q-ary separating codes the ratio between lower and upper bounds does not depend on q as $s \to \infty$. We leave it as an open question to close the gap between these lower and upper bounds.

Acknowledgements I.V. Vorobyev, N.A. Polyanskii and V.Yu. Shchukin have been supported in part by the Russian Science Foundation under Grant No. 14-50-00150.

References

1. Barg A., Blakley G.R., Kabatiansky G.A.: Digital fingerprinting codes: problem statements, constructions, identification of traitors. IEEE Trans. Inf. Theory **49**(5), 852–865 (2003).
2. Boneh D., Shaw J.: Collusion-secure fingerprinting for digital data. IEEE Trans. Inf. Theory **44**(5), 1897–1905 (1998).
3. Cohen G.D., Schaathun H.G.: Asymptotic overview on separating codes. Technical Report 248, Department of Informatics, University of Bergen, Norway (2003).
4. D'yachkov A.G., Macula A.J., Vilenkin P.A., Torney D.C.: Families of finite sets in which no intersection of ℓ sets is covered by the union of s others. J. Comb. Theory Ser. A **99**(2), 195–218 (2002).
5. D'yachkov A.G., Vorobyev I.V., Polyanskii N.A., Shchukin VYu: Bounds on the rate of disjunctive codes. Probl. Inf. Transm. **50**(1), 27–56 (2014).
6. D'yachkov A.G., Vorobyev I.V., Polyanskii N.A., Shchukin V.Yu.: Cover-free codes and separating system codes. In: Proceedings of the IEEE International Symposium on Information Theory (ISIT), pp. 2894–2898, IEEE, Hong Kong (2015).
7. Engel K.: Interval packing and covering in the Boolean lattice. Comb. Prob. Comput. **5**(4), 373–384 (1996).
8. Friedman A.D., Graham R.L., Ullman J.D.: Universal single transition time asynchronous state assignments. IEEE Trans. Comput. **18**(6), 541–547 (1969).
9. Lebedev V.S.: Asymptotic upper bound for the rate of (w, r) cover-free codes. Probl. Inf. Transm. **39**(4), 317–323 (2003).
10. Mago G.: Monotone functions in sequential circuits. IEEE Trans. Comput. **22**(10), 928–933 (1973).
11. Mitchell C.J., Piper F.C.: Key storage in secure networks. Discret. Appl. Math. **21**(3), 215–228 (1988).
12. Sagalovich Yu.L.: Cascade codes of automata states. Probl. Inf. Transm. **14**(2), 77–85 (1978).
13. Sagalovich Yu.L.: Completely separating systems. Probl. Inf. Transm. **18**(2), 140–146 (1982).
14. Sagalovich Yu.L.: Separating systems. Probl. Inf. Transm. **30**(2), 105–123 (1994).
15. Shangguan C., Wang X., Ge G., Miao Y.: New bounds for frameproof codes, preprint, 2014. arxiv:1411.5782v1
16. Staddon J.N., Stinson D.R., Wei R.: Combinatorial properties of frameproof and traceability codes. IEEE Trans. Inf. Theory **47**(3), 1042–1049 (2001).
17. Stinson D.R., Wei R., Chen K.: On generalized separating hash families. J. Comb. Theory Ser. A **115**(1), 105–120 (2008).