Combinatorial properties of codes with w-identifiable parents property*

Yu Xiong†
Institute of Applied Mathematics and Engineering Computations,
Hangzhou Dianzi University
Hangzhou 310018, P.R.China
Jun Ma‡
Department of Mathematics, Shanghai Jiaotong University,
Shanghai, 200240,P.R.China

Abstract. In this paper, we study the combinatorial properties of w-IPP (identifiable parents property) codes and give necessary and sufficient conditions for a code to be a w-IPP code. Furthermore, let $R(C) = \frac{1}{n}log_q|\mathcal{C}|$ denote the rate of the q-ary code C of length n, suppose $q \geq 3$ is a prime power, we prove that there exists a sequence of linear q-ary 2-IPP codes C_n of length n with $R(C_n) = \frac{1}{3}log_q\frac{q^3}{4q^2-6q+3}$.

MSC: 94A60, 05C65

Keywords: identifying parent property, fingerprinting, configuration.

1 Introduction

Motivated by an application in fingerprinting digital multimedia, the concept of a code with identifiable parent property was introduced by Hollmann et al. [2] and generalized by Staddon et al. [3].

Let $\mathcal{Q}$ be an alphabet of size q and $\mathcal{Q}^n$ denote the set of all n-tuples over $\mathcal{Q}$. Let $C \subseteq \mathcal{Q}^n$, $N = |C|$, then C is called a code of length n and size N, denoted an (N, n, q)-code. If $c = (c_1, c_2, \cdots, c_n) \in C$, then c is called a codeword of C.

*Project supported by National Natural Science Foundation of China under Grant No. 10471093.

†xiongyu@hdu.edu.cn
‡majun@math.sinica.edu.tw

Let C be an (N, n, q)-code. For $X \subseteq C$, we define the set of descendants of X, denoted $desc(X)$ by

$$desc(X) = \{d \in Q^n : d_i \in \{x_i : x \in X\}, 1 \leq i \leq n\}.$$

A set $X \subseteq C$ is said to be a parent set of a word $d \in Q^n$ if $d \in desc(X)$. For $d \in Q^n$, we write $\mathcal{H}_w(d, C)$ for the set of parent sets $X \subseteq C$ of d such that $|X| \leq w$.

Definition 1.1 *Let C be an (N, n, q)-code. If for each $d \in Q^n$, either $\mathcal{H}_w(d, C) = \emptyset$ or*

$$\bigcap_{X \in \mathcal{H}_w(d,C)} X \neq \emptyset,$$

then C is called a w-IPP code.

In [3], Staddon et al. mentioned an open problem: Is there a "tight" characterization of w-IPP codes for $w \geq 3$? In this paper, we shall answer this question and give a necessary and sufficient condition for a code to be a w-IPP code.

As usual, let $R(C) = \frac{1}{n} log_q |C|$ denote the rate of the q-ary code C of length n, $F_w(n, q) = \max\{|C| \mid C \text{ is a } q - \text{ary } w - \text{IPP code of length } n\}$ and $R_q(w) = \lim_{n \to \infty} \inf \frac{1}{n} \log_q F_w(n, q)$. For any $q \geq 3$, Hollmann et al. [2] proved

$$R_q(2) \geq \log_q (q/(4q^2 - 6q + 3)^{\frac{1}{3}}). \quad (1)$$

A.Barg et al. [1] strengthened this result by proving that there exists a sequence of linear 3-ary 2-IPP codes C_n of length n with $R(C_n) = \frac{1}{3} \log_3 \frac{9}{7}$. In this paper, let $q \geq 3$ be a prime power, we show that the bound in (1) can be achieved by a sequence of linear q-ary 2-IPP codes.

2 Necessary and sufficient conditions for a code to be a w-IPP code

Definition 2.1 *Suppose that C is a code, $X_1, X_2, \cdots, X_k$ are k distinct subcodes of C and $|X_i| = t$ for any $i = 1, 2, \cdots, k$. Let $H = \{X_1, X_2, \cdots, X_k\}$, if $\bigcap_{i=1}^{k} X_i = \emptyset$, then H is called (t, k)-configuration of C.*

Furthermore, Let H be (t, k)-configuration of C, if $\bigcap_{i=1, i \neq j}^{k} X_i \neq \emptyset$ for any $j = 1, 2, \cdots, k$, then H is called minimal (t, k)-configuration of C.

The following lemma tell us that any (t, k)-configuration must contain some minimal (t, k^*)-configurations $(2 \leq k^* \leq k)$.

Lemma 2.1 *Suppose C is a code, and $H = \{X_1, X_2, \cdots, X_k\}$ is a (t,k)-configuration of C. Then there exist positive integer $2 \leq k^* \leq k$ and $H^* \subseteq H$ such that H^* is a minimal (t,k^*)-configuration of C.*

Proof. Let $H_0 = H$, if H_0 satisfies $\bigcap_{i=1,i\neq j}^{k} X_i \neq \emptyset$ for any $j \in \{1,2,\cdots,k\}$, take $k^* = k$ and $H^* = H_0$, then H^* is a minimal (t,k^*)-configuration of C. Otherwise, there exists a $j \in \{1,2,\cdots,k\}$ such that $\bigcap_{i=1,i\neq j}^{k} X_i = \emptyset$. Without loss of generality, suppose $j = k$, then $\bigcap_{i=1}^{k-1} X_i = \emptyset$. Let $H_1 = \{X_1, \cdots, X_{k-1}\}$, then H_1 is (t,k^*)-configuration of C. Repeating the previous process, we obtain a sequence $H_0, H_1, \cdots, H_r$ where $0 \leq r \leq k-2$. Let $k^* = k - r$, $H^* = H_r$, then $|H^*| = k^*$ and H^* is a minimal (t,k^*)-configuration of C. □

To obtain the properties of minimal (t,k^*)-configurations of code C, we need the following definition.

Definition 2.2 *Suppose C be a code, $H = \{X_1, X_2, \cdots, X_k\}$ is a minimal (t,k^*)-configuration of C. For any $j \in \{1,2,\cdots,k\}$, let $b_j \in \bigcap_{i=1,i\neq j}^{k} X_i$, then the set $\{b_1, b_2, \cdots, b_k\}$ is called a frame of H.*

Lemma 2.2 *Suppose C is a code, $H = \{X_1, X_2, \cdots, X_k\}$ is a minimal (t,k^*)-configuration of C, $B = \{b_1, b_2, \cdots, b_k\}$ is a frame of H. Then*
(1)For any $j \in \{1,2,\cdots,k\}$, $B \backslash \{b_j\} \subseteq X_j$.
(2) $\left|\bigcup_{i=1}^{k} X_i\right| \leq \lfloor (\frac{1}{2}t+1)^2 \rfloor$.

Proof. (1) Let $j \in \{1,2,\cdots,k\}$. For any $m \neq j$, since $b_m \in \bigcap_{i=1,i\neq m}^{k} X_i$, we have $b_m \in X_j$. So, $B \backslash \{b_j\} \subseteq X_j$.

(2) By (1), we have $2 \leq k \leq t+1$. For any $j \in \{1,2,\cdots,k\}$, let $A_j = X_j \backslash (B \backslash \{b_j\})$, then $\bigcup_{j=1}^{k} X_j = B \cup (\bigcup_{j=1}^{k} A_j)$. Observe that $B \cap (\bigcup_{j=1}^{k} A_j) = \emptyset$. So, $|\bigcup_{j=1}^{k} X_j| = |B| + |(\bigcup_{j=1}^{k} A_j)|$.

Since $|B| + |(\bigcup_{j=1}^{k} A_j)| \leq |B| + \sum_{j=1}^{k} |A_j| = tk - k(k-2) = -k^2 + (t+2)k \leq (\frac{1}{2}t+1)^2$, we have $|\bigcup_{j=1}^{k} X_j| \leq \lfloor (\frac{1}{2}t+1)^2 \rfloor$. □

Now, suppose that $\mathcal{C}$ is a (N,n,q)-code, $N \geq w$, $d \in \mathcal{Q}^n$ and $\mathcal{H}_w(d,\mathcal{C}) \neq \emptyset$. For any $X \in \mathcal{H}_w(d,\mathcal{C})$, if $|X| < w$, since $N > w$, then there exists a subcode X' of $\mathcal{C}$ such that $X \subset X'$ and $|X'| = w$. So, let $\mathcal{H}_w^*(d,\mathcal{C}) = \{X' \mid X' \in \mathcal{H}_w(d,\mathcal{C}), |X'| = w\}$, then $\mathcal{H}_w^*(d,\mathcal{C}) \subseteq \mathcal{H}_w(d,\mathcal{C})$. Furthermore, we may obtain the following lemma.

Lemma 2.3 *Suppose $\mathcal{C}$ is an (N,n,q)-code, $N \geq w$, $d \in \mathcal{Q}^n$, $\mathcal{H}_w(d,\mathcal{C}) \neq \emptyset$. Then*

$$\bigcap_{X \in \mathcal{H}_w(d,\mathcal{C})} X = \bigcap_{X' \in \mathcal{H}_w^*(d,\mathcal{C})} X'.$$

Proof. For any $X \in \mathcal{H}_w(d,\mathcal{C})$, let $Y_X = \{X' \mid X \subseteq X' \subseteq \mathcal{C}\}$, then $\bigcap_{X' \in Y_X} X' = X$. So, $\bigcap_{X' \in \mathcal{H}_w^*(d,\mathcal{C})} X' = \bigcap_{X \in \mathcal{H}_w(d,\mathcal{C})} (\bigcap_{X' \in Y_X} X') = \bigcap_{X \in \mathcal{H}_w(d,\mathcal{C})} X$. □

Now, we prove the main theorem in this section.

Theorem 2.1 *Suppose $w \geq 2$, $\mathcal{C}$ is an (N,n,q)-code and $N \geq w$. The necessary and sufficient conditions for $\mathcal{C}$ to be a w-IPP code are for any $k \in \{2,3,\cdots,w+1\}$ and any minimal (w,k)-configuration $H = \{X_1, X_2, \cdots, X_k\}$ of $\mathcal{C}$, where $X_i = \{x_{i,1}, x_{i,2}, \cdots, x_{i,w}\}$, $x_{i,j} = (x_{i,j}^1, x_{i,j}^2, \cdots, x_{i,j}^n)$ for any $1 \leq i \leq k$ and $1 \leq j \leq w$, there exists a coordinate $l \in \{1,2,\cdots,n\}$ such that $\bigcap_{i=1}^{k} \{x_{i,1}^l, x_{i,2}^l, \cdots, x_{i,w}^l\} = \emptyset$.*

Proof. Suppose there exist $k \in \{2,3,\cdots,w+1\}$ and minimal (w,k)-configuration $H = \{X_1, X_2, \cdots, X_k\}$ of $\mathcal{C}$ such that $\bigcap_{i=1}^{k} \{x_{i,1}^l, x_{i,2}^l, \cdots, x_{i,w}^l\} \neq \emptyset$ for any coordinate $l = 1,2,\cdots,n$.

For any $l \in \{1,2,\cdots,n\}$, let $d_l \in \bigcap_{i=1}^{k} \{x_{i,1}^l, x_{i,2}^l, \cdots, x_{i,w}^l\}$ and $d = (d_1, d_2, \cdots, d_n)$. Clearly, $d \in desc(X_j)$ for any $1 \leq j \leq k$. But $\bigcap_{j=1}^{k} X_j = \emptyset$, a contradiction.

Conversely, suppose $\mathcal{C}$ isn't a w-IPP code. Then there exists $d \in \mathcal{Q}^n$ such that $\mathcal{H}_w(d,\mathcal{C}) \neq \emptyset$ and $\bigcap_{X \in \mathcal{H}_w(d,\mathcal{C})} X = \emptyset$. Since $N \geq w$, by Lemma 2.3, we have $\mathcal{H}_w^*(d,\mathcal{C}) \neq \emptyset$ and $\bigcap_{X \in \mathcal{H}_w^*(d,\mathcal{C})} X = \emptyset$. Observe that $\mathcal{H}_w^*(d,\mathcal{C})$ is a $(w, |\mathcal{H}_w^*(d,\mathcal{C})|)$-configuration of $\mathcal{C}$, by Lemmas 2.1 and 2.2, there exist $2 \leq k \leq w+1$ and $H \subseteq \mathcal{H}_w^*(d,\mathcal{C})$ such that H is a minimal (w,k)-configuration of $\mathcal{C}$.

Let $H = \{X_1, X_2, \cdots, X_k\}$, $X_i = \{x_{i,1}, x_{i,2}, \cdots, x_{i,w}\}$ and $x_{i,j} = (x_{i,j}^1, x_{i,j}^2, \cdots, x_{i,j}^n)$ for any $1 \leq i \leq k$ and $1 \leq j \leq w$. Then $d \in desc(X_i)$ for

$1 \leq i \leq k$. It follows that $\bigcap_{i=1}^{k} \{x_{i,1}^l, x_{i,2}^l, \cdots, x_{i,w}^l\} \neq \emptyset$ for any coordinate $l \in \{1, 2, \cdots, n\}$, a contradiction. □

From Theorem 2.1 and Lemma 2.2, we may obtain the following corollary.

Corollary 2.1 *[1] Suppose $C \subseteq Q^n$ is an (N, n, q)-code and $N \geq 2$. Then C is a 2-IPP code if and only if*

IPP1: a, b, c distinct in $C \Rightarrow a_i$, b_i , c_i distinct in Q for some i,

IPP2: $a, b, c, d \in C$ with $\{a, b\} \cap \{c, d\} = \emptyset \Rightarrow \{a_i, b_i\} \cap \{c_i, d_i\} = \emptyset$ for some i.

3 Linear 2-IPP codes.

For any $q \geq 3$, Hollmann et al. [2] proved

$$R_q(2) \geq \log_q(q/(4q^2 - 6q + 3)^{\frac{1}{3}}).$$

It follows that $R_3(2) \geq \frac{1}{3}\log_3 \frac{9}{7}$. A.Barg et al. [1] strengthened this result by proving that there exists a sequence of linear 3-ary 2-IPP codes C_n of length n with $R(C_n) = \frac{1}{3}\log_3 \frac{9}{7}$. In this section, let $q \geq 3$ is a prime power, we shall show that the bound can be achieved by a sequence of linear q-ary 2-IPP codes.

Theorem 3.1 *Suppose $q \geq 3$ is a prime power. Then there exists a sequence of linear q-ary 2-IPP codes C_n of length n with*

$$R(C_n) = \frac{1}{3} log_q \frac{q^3}{4q^2 - 6q + 3}.$$

Proof. Let C be a linear subspace of F_q^n . Consider the condition (IPP2) in Corollary 2.1. Suppose that $dim\ C = k$ and let G be a generator matrix of C, i.e., a $k \times n$ matrix whose rows form a basis of C as an F_q-linear space. Let g_1, g_2, $\cdots$, g_n be the columns of G. Any vector $c \in C$ has the form aG for some $a \in F_q^k$. Let c_1, $\cdots$, c_4 be some vectors in C. Since the 2-IPP property is translation invariant, we may suppose that $c_4 = 0$. Suppose that $c_i = a_i G$ for $i = 1, 2, 3$.

Case (a). a_1, a_2, a_3 are linearly independent. Complement a_1, a_2, a_3 to a basis and take the dual basis f_1, $\cdots$, f_k in F_q^k , so

$$a_i \bullet f_j = \begin{cases} 0 & if \quad i \neq j \\ 1 & if \quad i = j. \end{cases}$$

for $i = 1, 2, 3$ and $j = 1, 2, \cdots, k$. For any $m \in \{1, 2, \cdots, n\}$, let $h_{m,1}, h_{m,2}$, $\cdots$, $h_{m,k} \in F_q$ satisfy $\sum_{i=1}^{k} h_{m,i} f_i = g_m$. For any given $m = 1, 2, \cdots, n$, observe that $\{c_{1,m}, c_{2,m}\} \cap \{c_{3,m}, c_{4,m}\} = \emptyset$ if and only if $(h_{m,1}, h_{m,2}, h_{m,3})$ have one of the following forms:

$$\begin{array}{ll} (x, y, 0) & \textit{where} \quad x \neq 0, y \neq 0; \\ (x, y, z) & \textit{where} \quad z \neq 0, x \neq 0, y \neq 0, x \neq z, y \neq z. \end{array}$$

Hence the total number of favorable choices is $(q-1)^2 + (q-1)(q-2)^2$ out of q^3. This implies that the probability for a matrix G to be bad for a given linearly independent triple is $(1 - \frac{(q-1)^2+(q-1)(q-2)^2}{q^3})^n$. The number of triples is less than q^{3k}, so the probability that a given matrix spans a quadruple of vectors that violate the condition (IPP2) is bounded above by $q^{3k}(1 - \frac{(q-1)^2+(q-1)(q-2)^2}{q^3})^n$. Hence if $R = \frac{1}{3} log_q \frac{q^3}{4q^2-6q+3} - \varepsilon$ for any $\varepsilon > 0$, then there exists a favorable choice.

Case (b). Some of the vectors a_1, a_2, a_3 are linearly dependent. For instance, suppose that a_3 is spanned by a_1, a_2, and these two are not collinear. Let $a_3 = a_1 + a_2$. Choose a basis $f_1, \cdots, f_k$ in F_q^k such that

$$a_i \bullet f_j = \begin{cases} 0 & if \quad i \neq j \\ 1 & if \quad i = j. \end{cases}$$

for $i = 1, 2$ and $j = 1, 2, \cdots, k$. As above we count the number of unfavorable choices for g_m. Good choices for $(h_{m,1}, h_{m,2})$ are (x, y), where $x \neq 0$ and $y \neq 0$. Hence the fraction of bad choices of G is at most $q^{2k}(1 - \frac{(q-1)^2}{q^2})^n$, and this is less than $q^{3k}(1 - \frac{(q-1)^2+(q-1)(q-2)^2}{q^3})^n$. Other cases of dependence are dealt with analogously; none accounts for a fraction of bad matrices larger than in Case (a).

Now consider the condition (IPP1) in Corollary 2.1. Let c_1, c_2, c_3 be some vectors in C. Since the 2-IPP property is translation invariant we may assume that $c_3 = 0$. Suppose that $c_i = a_i G$ for $i = 1, 2$. There are the following two cases.

Case (a). a_1 and a_2 are linearly independent. Choose a basis $f_1, \cdots, f_k$ in F_q^k such that

$$a_i \bullet f_j = \begin{cases} 0 & if \quad i \neq j \\ 1 & if \quad i = j. \end{cases}$$

for $i = 1, 2$ and $j = 1, 2, \cdots, k$. For any $m \in \{1, 2, \cdots, n\}$, let $h_{m,1}, h_{m,2}$, $\cdots, h_{m,k} \in F_q$ satisfy $\sum_{i=1}^{k} h_{m,i} f_i = g_m$. For any given $m = 1, 2, \cdots, n$, observe that $c_{1,m} \neq c_{2,m}$, $c_{1,m} \neq 0$ and $c_{2,m} \neq 0$ if and only if $(h_{m,1}, h_{m,2})$ have one of the following forms:

$$(x, y) \quad \textit{where} \quad x \neq 0, y \neq 0, x \neq y.$$

Hence the total number of favorable choices is $(q-1)(q-2)$ out of q^2. This implies that the probability for a matrix G to be bad for a given linearly independent pair is $(1-\frac{(q-1)(q-2)}{q^2})^n$. The number of pairs is less than q^{2k}, so the probability that a given matrix spans a triple of vectors such that a_1 and a_2 are linearly independent, and such that they violate the condition (IPP1, is bounded above by $q^{2k}(1-\frac{(q-1)(q-2)}{q^2})^n$. Hence if $R=\frac{1}{2}log_q\frac{q^2}{3q-2}-\varepsilon$ for any $\varepsilon>0$, there exists a favorable choice.

Case (b). a_1 and a_2 are collinear, i.e., $a_1=\lambda a_2$. Choose a basis $f_1,\cdots,f_k$ in F_q^k so

$$a_1 \bullet f_j = \begin{cases} 1 & if \quad j=1 \\ 0 & if \quad j>1. \end{cases}$$

for $j=1,2,\cdots,k$. As above, good choices for $h_{m,1}$ are x, where $x\neq 0$. Hence the number of bad choices of G is at most $q^{2k}(\frac{1}{q})^n$, and this is less than $q^{2k}(1-\frac{(q-1)(q-2)}{q^2})^n$.

Thus, $\frac{1}{3}log_q\frac{q^3}{4q^2-6q+3}$ is the minimum of the achievable rates for conditions $IPP1$ and $IPP2$. Then we proved the bound (1) can be achieved by a sequence of linear q-ary 2-IPP codes. □

References

[1] A. Barg, G. Cohen, S. Encheva, G. Kabatiansky, G. Zémor, A hypergraph approach to the identifying parent property:the case of multiple parents. *SIAM J. Discrete Math. Vol. 14, No. 3(2001) pp. 423-431.*

[2] H.D.L. Hollmann, J.H. van Lint, J.-P. Linnartz, L.M.G.M. Tolhuizen, On codes with the identifiable parent property, *J Combin. Theory Ser.A 82(1998) 121-133.*

[3] J.N. Staddon, D.R.Stinson, T. Wei, Combinatorial properties of frameproof and traceability code, *IEEE Trans. Inform. Theory 47 (2001) 1024-1049.*