

Constructions for identifying codes

GEOFFREY EXOO ge@ginger.indstate.edu
Department of Mathematics and Computer Science, Indiana State
University, Terre Haute, IN 47809, USA

VILLE JUNNILA¹, TERO LAIHONEN² AND SANNA RANTO¹
viljun, terolai, samano@utu.fi
Department of Mathematics, University of Turku, 20014 Turku, FINLAND

Abstract. A nonempty set of words in a binary Hamming space \mathbf{F}^n is called an r -identifying code if for every word the set of codewords within distance r from it is unique and nonempty. The smallest possible cardinality of an r -identifying code is denoted by $M_r(n)$. In this paper, we consider questions closely related to the open problem whether $M_{t+r}(n+m) \leq M_t(m)M_r(n)$ is true. For example, we show results like $M_{1+r}(n+m) \leq 4M_1(m)M_r(n)$, which improve previously known bounds. We also obtain a result $M_1(n+1) \leq (2 + \varepsilon_n)M_1(n)$ where $\varepsilon_n \rightarrow 0$ when $n \rightarrow \infty$. This bound is related to the conjecture $M_1(n+1) \leq 2M_1(n)$. Moreover, we give constructions for the best known 1-identifying codes of certain lengths.

1 Introduction

Karpovsky, Chakrabarty and Levitin introduced identifying codes in [6] for locating malfunctioning processors in multiprocessor architectures. The research of identifying codes is also inspired by applications to sensor networks and alarm systems. Nowadays identifying codes are an actively studied topic of its own; the updated bibliography of identifying codes can be found from [7]. Identifying codes have been considered in many different graphs; in this paper we consider the binary Hamming spaces (i.e. binary hypercubes).

We denote by \mathbf{F}^n the binary Hamming space of dimension n . The (Hamming) *distance* between two vectors (called words) \mathbf{x} and \mathbf{y} in \mathbf{F}^n is denoted by $d(\mathbf{x}, \mathbf{y})$. The (Hamming) *weight* of a word \mathbf{x} , is denoted by $w(\mathbf{x})$. The (Hamming) *ball* of radius r centered at $\mathbf{x} \in \mathbf{F}^n$ is $B_r(\mathbf{x}) = \{\mathbf{y} \in \mathbf{F}^n \mid d(\mathbf{x}, \mathbf{y}) \leq r\}$.

A code of length n is a nonempty subset of \mathbf{F}^n . Let $C \subseteq \mathbf{F}^n$ be a code. The *I-set* of a word $\mathbf{x} \in \mathbf{F}^n$ (with respect to the code C) is defined to be

$$I_r(\mathbf{x}) = I_r(C; \mathbf{x}) = B_r(\mathbf{x}) \cap C.$$

¹Research supported by the Academy of Finland under grant 210280.

²Research supported by the Academy of Finland under grant 111940.

Definition 1 A code $C \subseteq \mathbf{F}^n$ is called an r -identifying if for all $\mathbf{x} \in \mathbf{F}^n$ $I_r(C; \mathbf{x}) \neq \emptyset$ and for all $\mathbf{y} \in \mathbf{F}^n$, $\mathbf{x} \neq \mathbf{y}$, we have

$$I_r(C; \mathbf{x}) \neq I_r(C; \mathbf{y}).$$

The definition of r -separating codes is similar to the identifying codes, but here we allow $I_r(\mathbf{x}) = \emptyset$ for one $\mathbf{x} \in \mathbf{F}^n$.

The *optimal*, that is, the smallest possible cardinality of an r -identifying code of length n is denoted by $M_r(n)$.

Notice that a code $C \subseteq \mathbf{F}^n$ is r -identifying if and only if for all $\mathbf{x}, \mathbf{y} \in \mathbf{F}^n$, $\mathbf{x} \neq \mathbf{y}$, we have $I_r(C; \mathbf{x}) \triangle I_r(C; \mathbf{y}) \neq \emptyset$ where the notation $A \triangle B$ denotes the symmetric difference of sets A and B , that is, $A \triangle B = (A \setminus B) \cup (B \setminus A)$.

A code $C \subseteq \mathbf{F}^n$ is called r -covering if for all $\mathbf{x} \in \mathbf{F}^n$ there is $\mathbf{c} \in C$ such that $d(\mathbf{x}, \mathbf{c}) \leq r$ (i.e., $|I_r(C; \mathbf{x})| \geq 1$). Moreover, if a code $C \subseteq \mathbf{F}^n$ has the property that for all $\mathbf{x} \in \mathbf{F}^n$ $|I_r(C; \mathbf{x})| \geq \mu$, then the code is called μ -fold r -covering. The optimal cardinality of an r -covering is denoted by $K(n, r)$. The vast topic of covering codes have been considered, for instance, in [3].

Let $C_1 \subseteq \mathbf{F}^n$ and $C_2 \subseteq \mathbf{F}^m$ be two codes, then their *direct sum*

$$C_1 \oplus C_2 = \{(\mathbf{a}, \mathbf{b}) \mid \mathbf{a} \in C_1, \mathbf{b} \in C_2\} \subseteq \mathbf{F}^{n+m}.$$

In [1], the question whether

$$M_{r+t}(n+m) \leq M_r(n)M_t(m) \tag{1}$$

holds is mentioned as an open problem. In [5] the result is proved for $r = t = 1$. In Section 2 of this paper, we consider the problems closely related to the conjecture (1) in a general case. In particular, we show that $M_{r+1}(n+m) \leq 4M_r(n)M_1(m)$ and also present some numerical improvements on known bounds on $M_r(n)$. In [1], it is also asked whether $M_1(n+1) \leq 2M_1(n)$ is true. In the last section, we show that $M_1(n+1) \leq (2 + \varepsilon_n)M_1(n)$ where $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

The proofs omitted in this paper are in [4].

2 New code constructions for r -identifying codes

In this section, we will present some direct sum constructions for $(r+t)$ -identifying codes. The motivation for this comes from the conjecture (1).

Lemma 1 Let $C \subseteq \mathbf{F}^n$ be an r -identifying code. Then for all $\mathbf{x} \in \mathbf{F}^n$ there exists $\mathbf{c} \in C$ such that $d(\mathbf{c}, \mathbf{x}) = r$ or $r+1$.

In the subsequent considerations we refer to the following condition for a given code C :

$$\forall \mathbf{x}, \mathbf{y} \in \mathbf{F}^n : I_t(C; \mathbf{x}) \setminus I_{t-1}(C; \mathbf{y}) \neq \emptyset. \quad (2)$$

We will use the following notations:

- The optimal cardinality of a t -identifying code satisfying the condition (2) is denoted by $\overline{M}_t(n)$.
- The optimal cardinality of a t -identifying code which is also $(t-1)$ -separating and satisfy the condition (2) is denoted by $\overline{M}_{t,t-1}(n)$.
- The optimal cardinality of a t -identifying code such that for every $\mathbf{x} \in \mathbf{F}^n$ there exists a codeword *exactly* at distance t from \mathbf{x} is denoted by $M'_t(n)$.
- We denote by $M''_1(n)$ the optimal cardinality of a 1-identifying and 2-fold 1-covering code. It is clear that $M''_1(n) \leq 2M_1(n)$.

Theorem 1 *We have*

$$M_{r+t}(n+m) \leq \begin{cases} M_r(n)\overline{M}_{t,t-1}(m), \\ M'_r(n)\overline{M}_t(m) \end{cases} \quad (3)$$

and

$$M_{r+1}(n+m) \leq M'_r(n)M''_1(m). \quad (4)$$

Moreover, $M'_r(n) \leq 2M_r(n)$. Especially,

$$M_{r+t}(n+m) \leq 2M_r(n)\overline{M}_t(n) \quad (5)$$

$$M_{r+1}(n+m) \leq 4M_r(n)M_1(m). \quad (6)$$

Proof. Let us first prove the inequalities (3). Let $C_1 \subseteq \mathbf{F}^n$ be an r -identifying code and $C_2 \subseteq \mathbf{F}^m$ be a t -identifying and $(t-1)$ -separating code satisfying the condition (2). We will first show that $C = C_1 \oplus C_2 \subseteq \mathbf{F}^{n+m}$ is an $(r+t)$ -identifying code. It is easy to see that C is an $(r+t)$ -covering code, this implies that $I_r(X) = \emptyset$ if and only if $X = \emptyset$. Therefore, in order to prove that C is $(r+t)$ -identifying, it is enough to show that $I_{r+t}(\mathbf{x}) \Delta I_{r+t}(\mathbf{y}) \neq \emptyset$ for all $\mathbf{x}, \mathbf{y} \in \mathbf{F}^{n+m}$ ($\mathbf{x} \neq \mathbf{y}$). Let $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$, $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2) \in \mathbf{F}^{n+m}$, where $\mathbf{x}_1, \mathbf{y}_1 \in \mathbf{F}^n$ and $\mathbf{x}_2, \mathbf{y}_2 \in \mathbf{F}^m$, moreover $\mathbf{x} \neq \mathbf{y}$.

1) Suppose first $\mathbf{x}_1 \neq \mathbf{y}_1$. Then there exists $\mathbf{c}_1 \in I_r(C_1; \mathbf{x}_1) \Delta I_r(C_1; \mathbf{y}_1)$. Without loss of generality we may assume that $\mathbf{c}_1 \in I_r(C_1; \mathbf{x}_1) \setminus I_r(C_1; \mathbf{y}_1)$. Since the code C_2 satisfies the condition (2), there exists a codeword $\mathbf{c}_2 \in C_2$ such that $\mathbf{c}_2 \in I_t(C_2; \mathbf{x}_2) \setminus I_{t-1}(C_2; \mathbf{y}_2)$. Hence, $(\mathbf{c}_1, \mathbf{c}_2) \in I_{r+t}(C; \mathbf{x}) \setminus I_{r+t}(C; \mathbf{y})$.

2) Suppose then $\mathbf{x}_1 = \mathbf{y}_1$. By Lemma 1, there exists $\mathbf{c}_1 \in C_1$ such that $d(\mathbf{c}_1, \mathbf{x}_1) = r$ or $r + 1$. Assume first that $d(\mathbf{c}_1, \mathbf{x}_1) = r$. Since C_2 is a t -identifying code and $\mathbf{x}_2 \neq \mathbf{y}_2$, there exists a codeword $\mathbf{c}_2 \in C_2$ such that $\mathbf{c}_2 \in I_t(\mathbf{x}_2) \triangle I_t(\mathbf{y}_2)$. Therefore, $(\mathbf{c}_1, \mathbf{c}_2) \in I_{r+t}(\mathbf{x}) \triangle I_{r+t}(\mathbf{y})$. Assume then that $d(\mathbf{c}_1, \mathbf{x}_1) = r + 1$. Since C_2 is also a $(t - 1)$ -separating code and $\mathbf{x}_2 \neq \mathbf{y}_2$, there exists a codeword $\mathbf{c}_2 \in C_2$ such that $\mathbf{c}_2 \in I_{t-1}(\mathbf{x}_2) \triangle I_{t-1}(\mathbf{y}_2)$. Hence, $(\mathbf{c}_1, \mathbf{c}_2) \in I_{r+t}(\mathbf{x}) \triangle I_{r+t}(\mathbf{y})$. Thus, we have proved that $C_1 \oplus C_2$ is an $(r + t)$ -identifying code.

Let $C_3 \subseteq \mathbf{F}^n$ be an r -identifying code such that for every $\mathbf{x} \in \mathbf{F}^n$ there exists a codeword exactly at distance r from it and $C_4 \subseteq \mathbf{F}^m$ a t -identifying code satisfying the condition (2). Showing that $C_3 \oplus C_4 \subseteq \mathbf{F}^{n+m}$ is an $(r + t)$ -identifying code is similar to the proof described above. However, in the second part of the proof we can assume that there always exists a codeword $\mathbf{c}_1 \in \mathbf{F}^n$ such that $d(\mathbf{x}_1, \mathbf{c}_1) = r$.

Let us now move on to the inequality (4). It is easy to see that 1-identifying and 2-fold 1-covering code satisfies the condition (2) for $t = 1$. Therefore, the result immediately follows from (3).

For the estimate $M'_r(n) \leq 2M_r(n)$, see [4]. \square

In [2, Theorem 3] it is proved that when $1 \leq t < m \leq r$ we have

$$M_{r+t}(n + m) \leq 2^m M_r(n). \quad (7)$$

Assume first $t = 1$. Since $C = \mathbf{F}^m \setminus \{1^m\}$ is clearly a 1-identifying and 0-separating code satisfying the condition (2), we have, by (3), that $M_{r+1}(n + m) \leq (2^m - 1)M_r(n)$. Using (6) we obtain further improvements to (7). Namely, we know that $M_1(m) \leq \frac{9}{2} \cdot \frac{2^m}{m+1} < 2^{m-2} - 1$ when $m \geq 18$ and, by the tables of [2], this also holds for $m \geq 8$.

In the next theorem we improve (7) using (5) when $t \geq 2$ and $m \geq 2t$. We give an upper bound for $\overline{M}_t(m)$ using a method inspired by Delsarte and Piret [3, p. 320].

Theorem 2 *Let $m \geq 2t$.*

$$M_{r+t}(n + m) \leq 2 \left[\frac{2^m}{\min\{\binom{m}{t}, 2\binom{m-1}{t}\}} 2m \ln 2 \right] M_r(n).$$

In what follows, we develop further the direct sum approach with the aid of k -locating-dominating codes. It is a class of codes introduced by Slater (see [8]) closely related to identifying codes; a code $C \subseteq \mathbf{F}^n$ is k -locating-dominating if $I_r(C; \mathbf{x})$ is nonempty and $I_r(C; \mathbf{x}) \neq I_r(C; \mathbf{y})$ for all non-codewords $\mathbf{x}, \mathbf{y} \in \mathbf{F}^n \setminus C$.

Theorem 3 Let $C_1 \subseteq \mathbf{F}^n$ be a 1-identifying code which is also a 2-fold 1-covering and has the property that it is k -locating-dominating for all $1 \leq k \leq r+1 \leq n-2$. Let $C_2 \subseteq \mathbf{F}^m$ be an r -identifying code. Then $C_1 \oplus C_2 \subseteq \mathbf{F}^{n+m}$ is an $(r+1)$ -identifying code.

The condition that the identifying code C_1 is a 2-fold 1-covering increases the cardinality only slightly (see [5]). The extra requirement that C_1 is also k -locating-dominating for $1 \leq k \leq n-2$ is not demanding cardinalitywise either. Indeed, the best 1-identifying 2-fold 1-covering codes which were found (Theorem 4), are immediately k -locating-dominating for all $1 \leq k \leq n-2$ as well.

Theorem 4 $M_1''(7) \leq 38$, $M_1''(8) \leq 70$, and $M_1''(10) \leq 249$.

It can also be checked that the best known 1-identifying and 2-fold 1-covering code of length 9 and of cardinality 128 [5] is k -locating-dominating for all $1 \leq k \leq 7$.

Corollary 1 $M_4(n) \leq 38M_3(n-7)$, $M_5(n) \leq 70M_4(n-8)$, $M_6(n) \leq 128M_5(n-9)$ and $M_7(n) \leq 249M_6(n-10)$.

The codes of Theorem 4 are also useful for bounding $M_1(n)$ from above. Namely, it has been proved in [5] that if a code $C \subseteq \mathbf{F}^n$ is 1-identifying and 2-fold 1-covering then the code $D = \{(\pi(\mathbf{u}), \mathbf{u}, \mathbf{u} + \mathbf{v}) \mid \mathbf{u} \in \mathbf{F}^n, \mathbf{v} \in C\} \subseteq \mathbf{F}^{2n+1}$ is 1-identifying and 2-fold 1-covering ($\pi(\cdot)$ is the parity check bit). Hence, we have the following theorem where the previous records are given in the parenthesis [2].

Theorem 5 $M_1(17) \leq 17920$ (18558) and $M_1(21) \leq 254976$ (262144).

A natural generalization of r -identifying codes are codes which identify sets of words, see [6]. A code $C \subseteq \mathbf{F}^n$ is called an $(r, \leq \ell)$ -identifying if for all $X, Y \subseteq \mathbf{F}^n$, $|X|, |Y| \leq \ell$, $X \neq Y$, we have

$$\bigcup_{\mathbf{x} \in X} I_r(C; \mathbf{x}) \neq \bigcup_{\mathbf{y} \in Y} I_r(C; \mathbf{y}).$$

The smallest cardinality of such codes in \mathbf{F}^n is denoted by $M_r^{(\leq \ell)}(n)$.

Theorem 6 Let r be a positive integer and suppose $\ell \geq r+3$. Let $C_1 \subseteq \mathbf{F}^{n_1}$ be a $(1, \leq \ell)$ -identifying code and $C_2 \subseteq \mathbf{F}^{n_2}$ be an $(r, \leq \ell)$ -identifying code. Then $C_1 \oplus C_2 \subseteq \mathbf{F}^{n_1+n_2}$ is an $(r+1, \leq \ell)$ -identifying code.

Corollary 2 When $r \geq 1$ and $\ell \geq r+3$ we have

$$M_{1+r}^{(\leq \ell)}(n+m) \leq M_1^{(\leq \ell)}(n)M_r^{(\leq \ell)}(m).$$

3 A direct sum of 1-identifying code and \mathbf{F}

In [1] it has been stated as an open problem whether $M_1(n+1) \leq 2M_1(n)$ holds, from there it also follows that $M_1(n+1) \leq 3M_1(n)$. The next theorem shows that $M_1(n+1) \leq (2 + \varepsilon_n)M_1(n)$ where $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

Theorem 7 *Assume $n \geq 2$. Then we have*

$$M_1(n+1) \leq \left(2 + \frac{1}{n+1}\right)M_1(n).$$

Proof. Let $C \subseteq \mathbf{F}^n$ be an optimal 1-identifying code attaining $M_1(n)$. Define

$$\begin{aligned} C_1 &= \{ \mathbf{x} \mid \mathbf{x} \in C, |I_1(\mathbf{x})| = 1 \} \text{ and} \\ N_1 &= \{ \mathbf{x} \mid \mathbf{x} \in \mathbf{F}^n, \mathbf{x} \notin C, |I_1(\mathbf{x})| = 1 \}. \end{aligned}$$

Clearly, $|C_1 \cup N_1| \leq M_1(n)$. Assume first $|C_1| \leq M_1(n)/(n+1)$. Let $D_1 = C \oplus \mathbf{F} \subseteq \mathbf{F}^{n+1}$. Denote $O_l = \mathbf{F}^n \oplus \{l\}$ where $l \in \mathbf{F}$. Assume $\mathbf{x} = (\mathbf{x}', a) \in \mathbf{F}^{n+1}$ with $\mathbf{x}' \in \mathbf{F}^n$ and $a \in \mathbf{F}$. Since C is 1-identifying, the set $I_1(D_1; \mathbf{x})$ can coincide only with the I -sets of words in O_{a+1} . If $|I_1(C; \mathbf{x}')| \geq 2$, then the word \mathbf{x} is uniquely identified by its I -set $I_1(D_1; \mathbf{x})$ since each word in O_{a+1} 1-covers a unique word in O_a . It can now be assumed that $|I_1(C; \mathbf{x}')| = 1$.

Assume $\mathbf{x}' \in N_1$, i.e. $I_1(C; \mathbf{x}') = \{\mathbf{x}' + \mathbf{e}\}$, where $\mathbf{e} \in \mathbf{F}^n$ is a word of weight 1. The only word in O_{a+1} which 1-covers the codeword $(\mathbf{x}' + \mathbf{e}, a)$ is the word $(\mathbf{x}' + \mathbf{e}, a+1)$. However, $|I_1(C; \mathbf{x}' + \mathbf{e})| \geq 2$ and therefore, as above, it can be said that \mathbf{x} is uniquely identified. If $\mathbf{x}' \in C_1$, then clearly, $I_1(D_1; (\mathbf{x}', a)) = I_1(D_1; (\mathbf{x}', a+1))$. But such a problematic case can be solved by adding one codeword to the code D_1 . Thus, we have the claim in this case.

Assume then $|C_1| > M_1(n)/(n+1)$. Let $\mathbf{z} \in \mathbf{F}^n$ be a word of weight 1. Consider then a code $D_2 \subseteq \mathbf{F}^{n+1}$ defined as

$$D_2 = (C \oplus \{0\}) \cup ((C + \mathbf{z}) \oplus \{1\}).$$

Assume $\mathbf{x} = (\mathbf{x}', a) \in \mathbf{F}^{n+1}$ with $\mathbf{x}' \in \mathbf{F}^n$ and $a \in \mathbf{F}$. If $|I_1(C; \mathbf{x}')| \geq 2$, then, as above, the word \mathbf{x} is uniquely identified by its I -set $I_1(D_2; \mathbf{x})$.

Assume now that $\mathbf{x}' \in C_1$, i.e. $I_1(C; \mathbf{x}') = \{\mathbf{x}'\}$. The only word in O_{a+1} which 1-covers the codeword (\mathbf{x}', a) is the word $(\mathbf{x}', a+1)$. However, $|I_1(D_2; (\mathbf{x}', a+1)) \cap O_{a+1}| \geq 2$ since $|I_1(D_2; (\mathbf{x}' + \mathbf{z}, a+1)) \cap O_{a+1}| = 1$ and the underlying code C is 1-identifying. Therefore, as before, it can be deduced that \mathbf{x} is uniquely identified by its I -set $I_1(D_2; \mathbf{x})$.

Assume then $\mathbf{x}' \in N_1$, i.e. $I_1(C; \mathbf{x}') = \{\mathbf{x}' + \mathbf{e}\}$, where $w(\mathbf{e}) = 1$. Again it suffices to consider the word $(\mathbf{x}' + \mathbf{e}, a+1)$. If $I_1(D_2; (\mathbf{x}', a)) = I_1(D_2; (\mathbf{x}' + \mathbf{e}, a+1))$

1)), then $I_1(D_2; (\mathbf{x}' + \mathbf{e}, a+1)) \cap O_{a+1} = \{(\mathbf{x}', a+1)\}$. Since $I_1(D_2; (\mathbf{x}', a)) \cap O_a = \{(\mathbf{x}' + \mathbf{e}, a)\}$, we have $d((\mathbf{x}', a), (\mathbf{x}' + \mathbf{e} + \mathbf{z}, a+1)) = 1$. Thus, $I_1(D_2; (\mathbf{x}', a)) = I_1(D_2; (\mathbf{x}' + \mathbf{e}, a+1))$ if and only if $\mathbf{e} = \mathbf{z}$. The code D_2 can clearly be made 1-identifying by adding a codeword to the set for each one of these problematic cases. Moreover, there exists a word $\mathbf{e}' \in \mathbf{F}^n$ of weight 1 such that

$$|\{\mathbf{x} \in \mathbf{F}^n \mid \mathbf{x} \notin C, I_1(\mathbf{x}) = \{\mathbf{x} + \mathbf{e}'\}\}| \leq \frac{|N_1|}{n}.$$

If we now choose $\mathbf{z} = \mathbf{e}'$, then we have, by the previous considerations, that

$$M_1(n+1) \leq 2M_1(n) + \frac{|N_1|}{n} \leq \left(2 + \frac{1}{n+1}\right)M_1(n).$$

□

References

- [1] U. Blass, I. Honkala, S. Litsyn, Bounds on identifying codes, *Discr. Math.* 241, 2001, 119-128.
- [2] I. Charon, G. Cohen, O. Hudry, A. Lobstein, New identifying codes in the binary Hamming space, *Europ. J. Combin.*, to appear.
- [3] G. Cohen, I. Honkala, S. Litsyn, A. Lobstein, *Covering Codes*, Elsevier, Amsterdam, 1997.
- [4] G. Exoo, V. Junnila, T. Laihonon, S. Ranto, Upper bounds for binary identifying codes, submitted.
- [5] G. Exoo, T. Laihonon, S. Ranto, Improved upper bounds on binary identifying codes, *IEEE Trans. Inform. Theory* 53, 2007, 4255-4260.
- [6] M. G. Karpovsky, K. Chakrabarty, L. B. Levitin, On a new class of codes for identifying vertices in graphs, *IEEE Trans. Inform. Theory* 44, 1998, 599-611.
- [7] A. Lobstein, Identifying and locating-dominating codes in graphs, a bibliography, Published electronically at <http://perso.enst.fr/~lobstein/debutBIBidetlocdom.pdf>.
- [8] P. J. Slater, Dominating and reference sets in a graph, *J. Math. Phys. Sci.* 22, 1988, 445-455.