Families of Optimal Codes for Strong Identification

Tero Laihonen
Sanna Ranto

Turku Centre for Computer Science

TUCS Technical Reports

No 418, August, 2001

Families of Optimal Codes for Strong Identification

Tero Laihonen

Department of Mathematics
University of Turku
FIN-20014 Turku, Finland
The research of this author was supported by the Academy of Finland under grant #46186.

Sanna Ranto

Turku Centre for Computer Science
Lemminkäisenkatu 14 A, 4th floor
FIN-20520 Turku, Finland

TUCS

Turku Centre for Computer Science
TUCS Technical Report No 418
August, 2001
ISBN 952-12-0866-X
ISSN 1239-1891

Abstract

Codes for strong identification are considered. The motivation for these codes comes from locating faulty processors in a multiprocessor system. Constructions and lower bounds on these codes are given. In particular, we provide two infinite families of optimal strongly identifying codes, which can locate up to two malfunctioning processors in a binary hypercube.

Keywords: Binary code, identifying code, strongly identifying code, optimal code, Hamming space

TUCS Research Group
Coding Theory Group

1 Introduction

Let F_2^n be the Cartesian product of n copies of the binary field F_2. The Hamming distance $d(x, y)$ between vectors x and y in F_2^n is the number of coordinates in which they differ. The Hamming weight $w(x)$ of x is defined as $d(x, 0)$.

The following scheme of finding the malfunctioning processors in a multiprocessor system was introduced by Karpovsky, Chakrabarty and Levitin in [10].

Suppose 2^n processors are labelled by distinct binary vectors of F_2^n and the processors are connected if and only if the Hamming distance between the corresponding labels equals one. A processor can check its neighbourhood within Hamming distance t. It reports "NO" if there is something wrong in its neighbourhood and "YES" otherwise. Assuming there are at most l malfunctioning processors, we want to choose a subset of processors (i.e. a code $C \subseteq F_2^n$) in such a way that based on their reports we can tell where the faulty processors are located. Of course, the smaller the code the better. Let us define the concept more formally.

As usual, we denote by $|X|$ the cardinality of a set X, $S_t(x) = S_t^n(x) = \{y \in F_2^n \mid d(x, y) = t\}$ and the Hamming sphere $B_t(x) = B_t^n(x) = \cup_{i=0}^t S_i^n(x)$. Let C be a subset of F_2^n, i.e., C is a code of length n. For any $X \subseteq F_2^n$ we define its "codeword neighbourhood" by

$$I_t(X) = I_t(C; X) = \left(\bigcup_{x \in X} B_t(x) \right) \cap C.$$

Definition 1. *Let t and l be non-negative integers. The code $C \subseteq F_2^n$ is called $(t, \leq l)$-identifying, if for all $X_1, X_2 \subseteq F_2^n$, $X_1 \neq X_2$, with $|X_1| \leq l$ and $|X_2| \leq l$ we have $I_t(X_1) \neq I_t(X_2)$.*

In the problem above we assume that also the malfunctioning processors can give the correct report. However, it is conceivable that faulty processors may send a wrong or right report [9]. This is the case we will be examining in this paper. We can analogously think that only the processors with the "NO" answer transmit the report and malfunctioning processors may send a report or be silent.

In order to find the malfunctioning processors in this case we require that C satisfies the following. Let for any distinct subsets X and Y of F_2^n ($|X|, |Y| \leq l$) the sets $I_t(X) \setminus S$ and $I_t(Y) \setminus T$ be different for all $S \subseteq X \cap C$ and $T \subseteq Y \cap C$. Then we can always distinguish between X and Y. The sets $I_t(X) \setminus S$ and $I_t(X) \setminus S'$ where $S, S' \subseteq X \cap C$ ($S \neq S'$) are automatically different from each other.

Definition 2. *Let $C \subseteq F_2^n$ and l and t non-negative integers. Define for $X \subseteq F_2^n$*

$$\mathcal{I}_t(X) = \{U \mid I_t(X) \setminus (X \cap C) \subseteq U \subseteq I_t(X)\}. \tag{1}$$

If for all $X_1, X_2 \subseteq F_2^n$, where $X_1 \neq X_2$ and $|X_1|, |X_2| \leq l$, we have $\mathcal{I}_t(X_1) \cap \mathcal{I}_t(X_2) = \emptyset$, then we say that C is a strongly $(t, \leq l)$-identifying code.

If we replace (1) by $\mathcal{I}_t(X) = \{I_t(X)\}$, we get Definition 1. Thus a strongly identifying code is an identifying code. In this paper we consider strongly $(1, \leq l)$-identifying codes when $l \geq 2$. Results for $l = 1$ can be found from [8]. We denote $I_t(\{x_1, \ldots, x_s\}) = I_t(x_1, \ldots, x_s)$ and $I_t'(y) = I_t(y) \setminus \{y\}$. The smallest cardinalities of a $(t, \leq l)$-identifying code and a strongly $(t, \leq l)$-identifying code of length n are denoted by $M_t^{(\leq l)}(n)$ and $M_t^{(\leq l)SID}(n)$, respectively. Usually we omit t from these notations if $t = 1$. A code attaining the smallest cardinality is called *optimal*. We say that x *covers* y, if $d(x, y) \leq 1$.

Clearly, F_2^n can be viewed as an undirected graph where all the binary words in F_2^n constitute the set of vertices and the edge set consists of all pairs of vertices connecting two words that are Hamming distance one apart. We call a sequence $v_0 e_1 v_1 e_2 \ldots e_n v_n$ of vertices v_i and edges $e_i = (v_{i-1}, v_i)$ a *cycle* if $v_i \neq v_j$ whenever $i \neq j$, except that $v_0 = v_n$.

2 General bounds

In the sequel, we shall often utilize the following well-known properties of a Hamming space (see, e.g, [3]). First of all, three Hamming spheres intersect in a unique point, if the intersection is nonempty. Secondly, due to the fact that moving by one edge changes the parity of the Hamming weight, there are no cycles of odd length in F_2^n. Moreover,

Lemma 1. *For $a, b \in F_2^n$ we have*

$$|B_1(a) \cap B_1(b)| = \begin{cases} n+1 & \text{if } a = b \\ 2 & \text{if } d(a,b) = 1 \text{ or } 2 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Theorem 1. *Let $l \geq 2$. Then*

$$M^{(\leq l)SID}(n) \geq \left\lceil (2l-1)\frac{2^n}{n} \right\rceil.$$

Proof. Let C be strongly $(1,\leq l)$-identifying. If $x \notin C$, then $|I(x)| \geq 2l-1$. Indeed, otherwise if $I(x) = \{c_1, \ldots, c_{2l-2}\}$ and x_i $(i = 1, \ldots, l-1)$ is the unique word different from x at distance one from both c_{2i-1} and c_{2i}, we have $I(x_1, \ldots, x_{l-1}) = I(x_1, \ldots, x_{l-1}, x)$, which is a contradiction. Obviously, fewer than $2l-2$ codewords in $I(x)$ is also impossible.

Assume then that $x \in C$. Suppose that $I(x) = \{c_1, \ldots, c_{2l-2}, x\}$ and define x_i as above for all $i = 1, \ldots, l-1$. Now $I(x_1, \ldots, x_{l-1}) = I(x_1, \ldots, x_{l-1}, x) \setminus \{x\}$ which is not allowed and hence $|I(x)| \geq 2l$.

Thus we obtain

$$|C|(n+1) \geq 2l|C| + (2l-1)(2^n - |C|)$$

which gives the claim. □

As we shall see, this lower bound can often be attained when $l = 2$. The *direct sum* of the codes $C_1 \subseteq F_2^{n_1}$ and $C_2 \subseteq F_2^{n_2}$ is

$$C_1 \oplus C_2 = \{(c_1, c_2) \mid c_1 \in C_1, c_2 \in C_2\} \subseteq F_2^{n_1+n_2}.$$

The *covering radius* $R(C)$ of a code $C \subseteq F_2^n$ is defined by

$$R(C) = \max_{x \in F_2^n} \min_{c \in C} d(x, c).$$

Evidently, $R(C_1 \oplus C_2) = R(C_1) + R(C_2)$.

Theorem 2. *If* $l \leq 2^{\lfloor \log_2(n+1) \rfloor - 1} - 1$, *then*

$$M^{(\leq l)SID}(n) \leq (2l+2)\frac{2^n}{2^{\lfloor \log_2(n+1) \rfloor}}.$$

Proof. We construct a strongly $(1,\leq l)$-identifying code C of length n provided that $l \leq 2^{\lfloor \log_2(n+1) \rfloor - 1} - 1$. Let H_r be the $[2^r-1, 2^r-1-r, 3]$ Hamming code, where $r = \lfloor \log_2(n+1) \rfloor$. Then $n = 2^r - 1 + m$ for some integer $m \geq 0$. By applying m times to H_r the direct sum with F_2, we obtain a code A with length n, dimension $n-r$ and covering radius one. Let the code C be a union of $2l+2$ different cosets of A.

Let X and Y be any distinct subsets of F_2^n of cardinality at most l and furthermore $S \subseteq X \cap C$ and $T \subseteq Y \cap C$. We show that always $I(Y) \setminus T \neq I(X) \setminus S$. Let us assume the contrary. Choose x such that $x \in X$ but $x \notin Y$. The word x is covered by at least $2l+2$ codewords of C. Denote $P = I'(x) \cap S$, i.e., P is the set of words in $I'(x)$ (it is convenient to work with $I'(x)$ instead of $I(x)$) that do not appear in $I(X) \setminus S$. Evidently, $P \subseteq X \setminus \{x\}$ and for all $c \in P$ we have $I(c) \cap I(x) \subseteq \{c, x\}$.

If $c \in P$ and $c \notin Y$, then the number of codewords in $(I(x) \cup I(c)) \setminus S$ needed to be covered by Y is at least $4l + 2 - |S| \geq 3l + 2$, since $|S| \leq l$. A word from Y cannot cover more than two words from $(I(x) \cup I(c)) \setminus \{x, c\}$, since $d(x, c) = 1$. Thus the set Y can cover at most $2|Y| + 2 = 2l + 2 < 3l + 2$ words of $(I(x) \cup I(c)) \setminus S$, which is a contradiction.

Hence we may assume that $P \subseteq Y$, and $P \subseteq S$ implies $P \subseteq T$. Now the number of codewords to be covered in $I'(x) \setminus P$ by Y is at least $|I'(x) \setminus P| \geq 2l + 1 - |P|$. On the other hand, because the words of P in Y can cover none of the words in $I'(x) \setminus P$ (and the rest of the words in Y can again cover at most two words), we can cover at most $2|Y \setminus P|$ words. Because $2l + 1 - |P| > 2l - 2|P| \geq 2|Y \setminus P|$, this is not enough, and thus we are done. Hence we can always distinguish between the sets $I(X) \setminus S$ and $I(Y) \setminus T$. $\square$

3 Optimal codes for strong $(1, \leq 2)$-identification

Let $C \subseteq F_2^n$ be $(1, \leq 2)$-identifying code. Then every word in F_2^n is covered by at least three codewords. This can be seen as follows. Obviously, every word $x \in F_2^n$ is covered by at least one codeword of C. If $x \in F_2^n$ is covered exactly by one codeword c, then $I(y) = I(x, y)$ for any $y \neq x$ such that $d(y, c) \leq 1$, a contradiction. If $I(x) = \{c_1, c_2\}$ where $c_1 \neq c_2$, then $I(y) = I(x, y)$ where $y = x + c_1 + c_2$ (notice that if $x = c_1$, then $y = c_2$ and vice versa), a contradiction. Therefore, we get the result below.

Lemma 2. *[7] If $C \subseteq F_2^n$ is $(1, \leq 2)$-identifying, then $|I(x)| \geq 3$ for all $x \in F_2^n$.*

The verification of the following fact can be found from [7].

Theorem 3. *Let C be a $(1, \leq 2)$-identifying. The direct sum of C and F_2 is also $(1, \leq 2)$-identifying.*

Theorem 4. *[8] A $(1, \leq 2)$-identifying code is strongly $(1, \leq 1)$-identifying.*

Proof. By Lemma 2 every word in F_2^n is covered by at least three codewords of a $(1, \leq 2)$-identifying code C. Thus $I(x)$ is unique for all $x \in F_2^n$. Furthermore, we have $I'(a) \neq I'(b)$ for all $a, b \in C$: especially, if $I'(a) = \{c_1, c_2\} = I'(b)$, then $I(c_1, a) = I(c_1, b)$ which is impossible. $\square$

Lemma 3. *Let $C \subseteq F_2^n$ be $(1, \leq 2)$-identifying and $a, b \in F_2^n$, $a \neq b$. Then*

$$|I(a, b) \setminus \{a, b\}| \geq \begin{cases} 2 & \text{if } d(a, b) = 1 \\ 3 & \text{if } d(a, b) = 2 \\ 4 & \text{if } d(a, b) \geq 3. \end{cases}$$

Proof. By Lemma 2, $|I(x)| \geq 3$ for all $x \in F_2^n$. By virtue of (2) the set $I(a,b)$ contains at least four elements. This gives the claim for $d(a,b) = 1$. If $|I(a,b) \setminus \{a,b\}| = 2$ and $d(a,b) = 2$, then $I(a,b) = \{a,b,c_1,c_2\}$ where $B_1(a) \cap B_1(b) = \{c_1,c_2\}$ for some $c_1, c_2 \in C$. However, now $I(a,c_1) = I(b,c_1)$, which is forbidden in C. Thus $|I(a,b) \setminus \{a,b\}| \geq 3$ for $d(a,b) = 2$. If $d(a,b) \geq 3$, then $B_1(a) \cap B_1(b) = \emptyset$ and hence we get the last claim. □

The following theorem turns out to be very useful when we construct (optimal) strongly $(1, \leq 2)$-identifying codes.

Theorem 5. *If C is a $(1, \leq 2)$-identifying code, then $D = C \oplus F_2$ is strongly $(1, \leq 2)$-identifying.*

Proof. By Theorem 3 we know that D is $(1, \leq 2)$-identifying and by Theorem 4 we know that hence D is strongly $(1, \leq 1)$-identifying. Thus to prove the claim it suffices to check the following two sets of inequalities for all x, y, z and w in F_2^{n+1} and for all sets J where $I'(a) \subseteq J(a) \subseteq I(a)$ and $I(a,b) \setminus \{a,b\} \subseteq J(a,b) \subseteq I(a,b)$: the first set is

$$J(x) \neq J(z,w) \tag{3}$$

where $z \neq w$ and $J(x) \neq I(x)$ or $J(z,w) \neq I(z,w)$, and the second set is

$$J(x,y) \neq J(z,w) \tag{4}$$

where $\{x,y\} \neq \{z,w\}$ and $J(x,y) \neq I(x,y)$ or $J(z,w) \neq I(z,w)$.

By Lemma 2 we have $|I(C;x)| \geq 3$ for all $x \in F_2^n$ (throughout this proof we write $I(C;X)$, but the set $I(D;X)$ is written $I(X)$ for short) and therefore also $|I(y)| \geq 3$ for all $y \in F_2^{n+1}$ and, moreover, $|I(y)| \geq 4$ if $y \in D$. Thus $|I'(y)| \geq 3$ for all $y \in F_2^{n+1}$.

Step 1: Let us first look at the inequalities (3). Either $x \neq z$ or $x \neq w$, say $x \neq z$. By (2), $|B_1(x) \cap B_1(z)| \leq 2$. If $d(x,z) \neq 2$, then there are at least two codewords in $I'(z)$ which are not in $I(x)$ and only one of them can be removed from $J(z,w)$. If $d(x,z) = 2$, then there can be only one such codeword and it can be removed from $I(z,w)$ if it is w. However, then the words in $I'(w)$ cannot be in $I(x)$, since $d(x,w) = 3$. This shows that (3) is satisfied.

Step 2: Consider next the inequalities (4). We denote by z' the word obtained by puncturing the last coordinate of $z \in F_2^{n+1}$. In the sequel we will often use the fact that $I(a) \cap (F_2^n \oplus \{1\})$, where $a = a'0 \in F_2^{n+1}$, contains the unique word $a'1$ if $a \in D$ and otherwise it is empty. Denote $L_0 := F_2^n \oplus \{0\}$ and $L_1 := F_2^n \oplus \{1\}$.

Case 1: Let $x, y, z, w \in L_0$. In the inequalities (4) we may assume that x is removed from $I(x, y)$. Thus $x \in D$. Consequently, by the fact above $x \in \{z, w\}$, say $x = z$. If also $y \in D$, then $\{x, y\} = \{z, w\}$. Similarly, we can assume that $w \notin D$. Let then $y \notin D$ and thus y cannot be removed from $I(x, y)$. Hence it suffices to verify that $I(x, y) \setminus \{x\} \neq I(z, w)$ and $I(x, y) \setminus \{x\} \neq I(z, w) \setminus \{z\}$. The first is immediately clear, since $x \in I(z, w)$. The second follows, because D is $(1, \leq 2)$-identifying, we have $I(x, y) \neq I(z, w)$. This proves (4) in this case.

Case 2: Let $x, y \in L_0$ and $z, w \in L_1$. Evidently, $|I(z, w) \cap L_0| \leq 2$. Therefore, by Lemma 3, we only need to examine the case, where $d(x, y) = 1$, both x and y are removed from $I(x, y)$ and $|(I'(x) \cap L_0) \setminus I(y)| = |(I'(y) \cap L_0) \setminus I(x)| = 1$. By symmetry, we can assume that the analogous premises hold for z and w as well, and thus only the inequality $I(x, y) \setminus \{x, y\} \neq I(z, w) \setminus \{z, w\}$ is left to be verified. Let $(I(x, y) \cap L_0) \setminus \{x, y\} = \{c_1, c_2\}$ for some $c_1, c_2 \in D$. If the inequality fails, we must have $z = c_1'1$ and $w = c_2'1$. Similarly, $(I(z, w) \cap L_1) \setminus \{z, w\} = \{x'1, y'1\}$. But this is a contradiction, since now $I(C; \{x', y'\}) = I(C; \{z', w'\})$ although $\{x', y'\} \neq \{z', w'\}$.

Case 3: Let $x, y, z \in L_0$ and $w \in L_1$. Since $|I(x, y) \cap L_1| \leq 2$, we only need to check the situation where $w \in D$ is removed from $I(z, w)$ and $|I'(w) \cap L_1| = 2$. Let $I'(w) \cap L_1 = \{c_1, c_2\}$, $c_1, c_2 \in D$. We may assume $c_1' = x'$ and $c_2' = y'$, and furthermore we can assume that x and y cover the words in $I'(z) \cap L_0$ (if $z \in D$, then it suffices to examine the cases $z \in \{x, y, w'0\}$) and z covers the words in $(I(x, y) \cap L_0) \setminus \{x, y, w'0\}$, otherwise we are done. Hence x' and y' cover the words in $I(C; z')$, and thus (4) hold, because if not, then $I(C; \{x', y'\}) = I(C; \{z', w'\})$ although $\{x', y'\} \neq \{z', w'\}$.

Case 4: Let finally $x, z \in L_0$ and $y, w \in L_1$. If $d(x, z) \geq 3$ or $d(y, w) \geq 3$, then by (2) the inequalities (4) obviously hold. Suppose that this is not the case. Assume first that $d(x, z) = 0$. Clearly, $d(y, w) \neq 0$ by $\{x, y\} \neq \{z, w\}$. If $d(y, w) = 1$, then according to (2), $|(I'(w) \cap L_1) \setminus I(y)| \geq 1$ and $|(I'(y) \cap L_1) \setminus I(w)| \geq 1$. These codewords cannot all be covered by $x = z$. Let next $d(y, w) = 2$. By the previous lemma, we can assume that there is at least one codeword, say c, such that $c \in (I'(w) \cap L_1) \setminus I(y)$. We may suppose $x' = z' = c'$. If there is a codeword in $(I'(y) \cap L_1) \setminus I(w)$, we are done, because $x = z$ cannot cover two distinct words in L_1. If $(I'(y) \cap L_1) \setminus I(w) = \emptyset$, then $y \in D$ (and it is removed from $I(x, y)$) and thus we must check the case $y'0 \in B_1(z) \setminus \{z\}$ ($y' \neq z'$, because $y' = z'$ implies $d(y, w) = 1$), but then $B_1(w') \cap B_1(y')$ contains x' and the two words of $B_1(w) \cap B_1(y)$, of which the last coordinates are deleted. This is impossible by (2).

Suppose next that $d(x, z) = 1$. The cases $1 \leq d(y, w) \leq 2$ need to be investigated. Let $d(y, w) = 1$. Using the fact that $|I(a) \cap L_i| \geq 3$ for all $a \in L_i$ $(i = 0, 1)$, this situation is impossible (at least one of the words x, y, z

or w is removed). Suppose next that $d(y,w)=2$. Let $a \in B_1(y) \cap B_1(w)$. By Lemma 3, it suffices to study the case where $(I'(x) \cap L_0) \setminus I(z) = \{w'0\}$ and $(I'(z) \cap L_0) \setminus I(x) = \{y'0\}$. Now there is a cycle of odd length ($x' \to z' \to y' \to a' \to w' \to x'$, here "$\to$" marks edges in the natural way), which is not possible in a Hamming space.

Let finally $d(x,z)=2$. By the discussion above, $d(y,w)=2$ remains to be explored. By Lemma 3, we can assume that $(I'(w) \cap L_1) \setminus I(y) \neq \emptyset$. If $|(I'(w) \cap L_1) \setminus I(y)| \geq 2$, we are done. It suffices to check the inequalities (4) with the choice $(I'(w) \cap L_1) \setminus I(y) = \{x'1\}$. Hence $d(x',w')=1$. Next we show that we may assume $d(y',z')=1$ as well. If $y \notin D$, then $|(I(y) \cap L_1) \setminus I(w)| \geq 1$ and it is enough to assume $(I(y) \cap L_1) \setminus I(w) = \{z'1\}$. This implies $d(z',y')=1$. Let $y \in D$. Obviously, $y' \neq x'$, since $d(w,y)=2$. If $y'0 \in B_1(x) \setminus \{x\}$, then $|B_1(y') \cap B_1(w')| \geq 3$, which is not possible by (2). Clearly, $y' \neq z'$ due to the fact that $y'=z'$ would create a cycle of odd length ($y' \to a \to x' \to w' \to b \to y'$) where $a \in B_1(x') \cap B_1(y')$ and $b \in B_1(w') \cap B_1(y')$). Thus, if $y \in D$, it suffices to check the case where $\{y'0\} = (I'(z) \cap L_0) \setminus I(x)$. Consequently, $d(z',y')=1$. If $|(I'(x) \cap L_0) \setminus I(z)| \geq 1$, then it is enough to examine the case $(I'(x) \cap L_0) \setminus I(z) = \{w'0\}$ (similarly, $(I'(y) \cap L_1) \setminus I(w) = \{z'1\}$ and $(I'(z) \cap L_0) \setminus I(x) = \{y'0\}$, if the sets on the left-hand sides are nonempty). Therefore, we always get $I(C;\{x',y'\}) = I(C;\{z',w'\})$ although $\{x',y'\} \neq \{z',w'\}$. Hence the inequalities (4) are always satisfied and this completes the proof of the assertion. □

Corollary 1. $M^{(\leq 2)SID}(n) \leq 2M^{(\leq 2)}(n-1)$.

We are now in a position to give two infinite families of optimal codes.

Corollary 2.

$$\text{For } k \geq 1: \quad M_1^{(\leq 2)SID}(3 \cdot 2^k) = 2^{3 \cdot 2^k - k}.$$
$$\text{For } k \geq 3: \quad M_1^{(\leq 2)SID}(2^k) = 3 \cdot 2^{2^k - k}.$$

Proof. From [11] we know that $M_1^{(\leq 2)}(n) = 2^{3 \cdot 2^k - k - 1}$, if $n = 3 \cdot 2^k - 1$ ($k \geq 1$), and $M_1^{(\leq 2)}(n) = 3 \cdot 2^{2^k - k - 1}$, if $n = 2^k - 1$ ($k \geq 3$). Combining this with Corollary 1 and the lower bound from Theorem 1 we obtain the equations. □

No infinite family of optimal strongly $(1, \leq 1)$-identifying codes is known.

Corollary 3. *If C is strongly $(1, \leq 2)$-identifying, then the direct sum with F_2 is as well.*

Before presenting one more optimality result, we show a construction which yields a strongly $(1, \leq 2)$-identifying code of length $2n+1$ from a strongly $(1, \leq 2)$-identifying code of length n.

Theorem 6. *Let C be a strongly $(1, \leq 2)$-identifying code of length n. The code*

$$C' = \{(\pi(u), u, u+c) \mid u \in F_2^n, c \in C\},$$

where $\pi(u)$ denotes the parity check bit of u, is a strongly $(1, \leq 2)$-identifying code of length $2n+1$.

Proof. By Theorem 3 C' is $(1, \leq 2)$-identifying and by Theorem 4 it is hence strongly $(1, \leq 1)$ identifying. Therefore, it is again enough to confirm the inequalities in (3) and (4).

Let us divide the words of F_2^{2n+1} into two classes by their first bit and consider the codewords which cover a word in each class. Let $x = (a, u, u+v) \in F_2^{2n+1}$.

I If $a = \pi(u)$ then $I(x) = \{(\pi(u), u, u+c) \mid c \in C, d(c, v) \leq 1\}$.

II If $a \neq \pi(u)$ then $I(x) = A \cup \{(a, u', u+v) \mid d(u', u) = 1, \exists c \in C : u+v = u'+c\}$. Here $A = \{(\pi(u), u, u+v)\}$ if $v \in C$, and $A = \emptyset$ if $v \notin C$.

Hence in both classes we are interested in codewords $c \in C$ such that $d(c, v) \leq 1$. Namely, in the class II the properties $d(u', u) \leq 1$ and $u+v = u'+c$ imply that also $d(v, c) \leq 1$. If $I(x) = \{(b_i, s_i, t_i) \mid i = 1, 2, \dots, k\}$, then in both cases $I(C; v) = \{s_i + t_i \mid i = 1, 2, \dots, k\}$. If $x \in C'$ and it is removed from $I(x)$, then v is removed from $I(C; v)$. Notice that if $I(x, y) \setminus \{x, y\} = \{(b_i, s_i, t_i)\} \mid i = 1, 2, \dots, k\}$, where $x = (a_1, u_1, u_1 + v_1)$ and $y = (a_2, u_2, u_2 + v_2)$, then $I(C; \{v_1, v_2\}) \setminus \{v_1, v_2\} \subseteq \{s_i + t_i \mid i = 1, 2, \dots, k\} \subseteq I(C; \{v_1, v_2\})$. Similarly, if $I(x, y) \setminus \{x\} = \{(b_i, s_i, t_i) \mid i = 1, 2, \dots k\}$, then $I(C; \{v_1, v_2\}) \setminus \{v_1\} \subseteq \{s_i + t_i \mid i = 1, 2, \dots, k\} \subseteq I(C; \{v_1, v_2\})$.

By the proof of Theorem 1 and [3, Theorems 3.4.3 and 14.4.3], $|I'(e)| \geq 3$ for all $e \in F_2^{2n+1}$. Hence we can prove the inequalities (3) using exactly the same argument as in the step 1 of the proof of Theorem 5.

Suppose one the inequalities (4) does not hold. Denote $x = (a_1, u_1, u_1 + v_1)$, $y = (a_2, u_2, u_2 + v_2)$, $z = (a_3, u_3, u_3 + v_3)$ and $w = (a_4, u_4, u_4 + v_4)$. Since C is a strongly $(1, \leq 2)$-identifying code we must have $\{v_1, v_2\} = \{v_3, v_4\}$ by the previous discussion.

It is enough to study the case, where all words x, y, z and w are in the class I, and the case where x and z are in the first class and y and w in the second one. We will show next that other cases need not to be considered.

Let x and y be in the first class and w in the second class. If $J(z,w)$ contains at least three elements of $I(w)$, then $J(z,w)$ has at least three codewords with different $n+1$ first bits and $J(x,y)$ can only provide two such words. If $J(z,w) = I(z,w) \setminus \{z\}$, where $z \in I(w)$, then $J(z,w)$ might only contain two words of $I(w)$. Denote $I(w) = \{z,a,b\}$ and $\{d,e,f\} \subseteq I'(z)$. Because w is in the class II, both a and b cannot belong to $I(x)$ or $I(y)$, say $a \in I(x)$. Furthermore, at least two elements of $I'(z)$ belong to $I(x)$ or $I(y)$. Let $\{d,e\} \subseteq I(x)$. Now $d \neq x \neq e$ (recall that there cannot be odd cycles), so if $x \neq a$, there exists an odd cycle ($w \to a \to x \to d \to z \to w$), and if $x = a$, then $B_1(x) \cap B_1(z) = \{w,d,e\}$, which is impossible by (2). Therefore, these cases where x and y are in the class I and w or both w and z lie in the second class need not to be examined.

Let then x belong to the class I and y, z and w to the second. Since $J(x,y)$ contains at least three codewords with different n last bits and $J(z,w)$ can only provide two such words, this situation is also impossible.

Because there are no codewords in the class II, it is enough to study the case, where x, y, z and w belong to the class I, and the case where x and z belong to the class I and y and w to the class II.

Let first all x, y, z and w lie in the first class. If $d(x,y) \geq 2$ and $d(z,w) \geq 2$, then $I'(x) \cup I'(y) \subseteq J(x,y)$ and by $\{x,y\} \neq \{z,w\}$ we obtain $u_1 = u_2 = u_3 = u_4$. If $d(x,y) = 1$ or $d(z,w) = 1$, then $u_1 = u_2 = u_3 = u_4$ also holds due to the fact that x, y, z and w are in the class I. From $u_1 = u_2 = u_3 = u_4$ it follows that $\{x,y\} = \{z,w\}$, a contradiction.

Assume next that x and y belong to the first class and y and w lie in the second. Because y and w cannot be codewords, it suffices to verify only the following inequalities of (4): $I(x,y) \setminus \{x\} \neq I(z,w) \setminus \{z\}$ and $I(x,y) \setminus \{x\} \neq I(z,w)$.

Suppose $I(x,y) \setminus \{x\} = I(z,w) \setminus \{z\}$. We have $x \neq z$, since otherwise $I(x,y) = I(z,w)$, which is impossible due to the fact that C' is $(1, \leq 2)$-identifying. Now we can assume $I'(x) \cap I'(z) \neq \emptyset$ and hence $u_1 = u_3$. Because $|I'(e)| \geq 3$ for all $e \in F_2^{2n+1}$, we obtain $|I'(x) \cap I(w)| = 1$ and $|I'(z) \cap I(y)| = 1$ (the number of elements in the intersections cannot be greater than one, since x and w and also y and z belong to different classes). The words in these two intersections must be different, since $x \neq z$. Since words y and w belong to the class II and we may assume that $I(y) \cap I(w) \neq \emptyset$, the last n bits of the words in $I(y)$ and $I(w)$ are the same. On the other hand, the words in the intersections $I'(x) \cap I(w)$ and $I'(z) \cap I(y)$ must also end with those n bits, but now $|I'(x) \cap I'(z)| \geq 3$, since $u_1 = u_3$, and thus $x = z$, a contradiction.

Assume finally $I(x,y) \setminus \{x\} = I(z,w)$. Evidently, $x \in C'$ and $x \neq z$. If z is a codeword, then in $I(z)$ there are at least two codewords which are not in $I'(x)$, these two words cannot be both in $I(y)$ either (y and z are in different classes), so $z \notin C'$. Now $I(x,y) \setminus \{x\} = I(z,w) \setminus \{z\}$, where $z \notin C'$ but belongs to the class I. This is impossible by the previous case. $\square$

There does not exist a strongly $(1, \leq 2)$-identifying code of length less than five. Indeed, for length four we notice that there cannot be a strongly $(1, \leq 2)$-identifying code, since always $I(0000, 1111) \setminus \{0000, 1111\} = I(0011, 1100) \setminus \{0011, 1100\}$ regardless of the code. From [7] it is known that there are no $(1, \leq 2)$-identifying code of length less than four.

Theorem 7. $M^{(\leq 2)SID}(5) = 22$.

Proof. By [7], $M^{(\leq 2)}(4) = 11$, and thus we obtain the upper bound using Corollary 1.

We prove the lower bound by using the fact observed in the proof of Theorem 1 that $|I'(x)| \geq 3$ for all $x \in F_2^5$. Suppose $M^{(\leq 2)SID}(5) \leq 21$. By [3, p. 383] we need at least 22 codewords to cover each word in F_2^5 at least four times. Hence we know that there is a noncodeword which is covered by exactly three codewords. Let C be a code attaining the bound $M^{(\leq 2)SID}(5)$. Without loss of generality, assume $00000 \notin C$, and $I(00000) = \{10000, 01000, 00100\}$. Each word of weight one must be covered by at least three codewords of weight two. Each codeword of weight two covers two words of weight one. Thus we need at least $\lceil 3 \cdot 5/2 \rceil = 8$ codewords of weight two.

If $11111 \notin C$ this is a symmetric situation, so all in all there are now at least 22 codewords.

Suppose $11111 \in C$. Now again there are at least three codewords of weight four. The word 00011 must be covered by three codewords of weight three, those can only be $\{10011, 01011, 00111\} =: T$. The words of the set $A := S_1^3(000) \oplus S_1^2(00)$ must be covered by at least two codewords of weight three. Denote by $G_a = \{a_1 a_2 a_3\} \oplus S_1^2(00)$, where $a = (a_1, a_2, a_3) \in S_1^3(000)$. The words of weight three always cover exactly two words of A, except that 11100 covers none. Moreover, a word in $S_3^5(00000) \setminus (\{11100\} \cup T)$ covers a word both in G_a and G_b for some a and b $(a \neq b)$. Let $c_1 \in S_3^5(00000)$ cover a word in G_a and G_b and, moreover, $c_2 \in S_3^5(00000)$ cover the other word in G_a and a word in G_d, where $b \neq a \neq d$ and $b \neq d$. The distance between the yet uncovered (by others than the words in T) words of G_b and G_d is four. Thus we need at least seven codewords of weight three in C. Again there are at least 22 codewords. $\square$

n	$M^{(\leq 2)}(n)$	$M^{(\leq 2)SID}(n)$
4	11	-
5	16	a 22 d
6	30–32	b 32 c
7	48	b 55–64 e
8	90–96	b 96 c
9	154–176	b 171–192 e
10	289–352	b 308–352 d
11	512	b 559–704 f,e
12	972-1024	b 1024 c
13	1756–2048	b 1891–2048 f,e
14	3356–4096	b 3511–4096 e
15	6144	b 6554–8192 f,e
16	11566–12288	b 12288 c

Table 1: Bounds on regular and strong $(1, \leq 2)$-identification. All the bounds on $M^{(\leq 2)}(n)$ are from [7, 11].

Key to Table 1	
a	Theorem 7
b	Theorem 1
c	Corollary 2
d	Corollary 1
e	Corollary 3
f	Theorem 6

Acknowledgements

The authors wish to thank Iiro Honkala for many inspiring discussions and comments which helped to improve the paper.

References

[1] U. Blass, I. Honkala and S. Litsyn, Bounds on identifying codes, Discrete Math., to appear.

[2] U. Blass, I. Honkala and S. Litsyn, On binary codes for identification, J. Combin. Des. 8 (2000) 151-156.

[3] G. Cohen, I. Honkala, S. Litsyn and A. Lobstein, Covering Codes (Elsevier, Amsterdam, 1997).

[4] G. Cohen, I. Honkala, A. Lobstein and G. Zémor, On identifying codes, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Proceedings of the DIMACS Workshop on Codes and Association Schemes, November 9–12, 1999, to appear.

[5] G. Exoo, Computational results on identifying t-codes, preprint.

[6] I. Honkala, On the identifying radius of codes, Proceedings of the Seventh Nordic Combinatorial Conference, Turku, pp. 39–43, 1999.

[7] I. Honkala, T. Laihonen and S. Ranto, On codes identifying sets of vertices in Hamming spaces, Des. Codes Cryptogr., to appear.

[8] I. Honkala, T. Laihonen and S. Ranto, On strongly identifying codes, Discrete Math., submitted.

[9] I. Honkala, T. Laihonen and S. Ranto, Codes for strong identification, Electronic Notes in Discrete Mathematics, to appear.

[10] M.G. Karpovsky, K. Chakrabarty and L. B. Levitin, On a new class of codes for identifying vertices in graphs, IEEE Trans. Inform. Theory 44 (1998) 599–611.

[11] S. Ranto, I. Honkala and T. Laihonen, Two families of optimal identifying codes in binary Hamming spaces, IEEE Trans. Inform. Theory, submitted.

Turku Centre for Computer Science
Lemminkäisenkatu 14
FIN-20520 Turku
Finland

http://www.tucs.abo.fi

University of Turku
- **Department of Mathematical Sciences**

Åbo Akademi University
- **Department of Computer Science**
- **Institute for Advanced Management Systems Research**

Turku School of Economics and Business Administration
- **Institute of Information Systems Science**