On Strongly Identifying Codes

Iiro Honkala
Tero Laihonen
Sanna Ranto

Turku Centre for Computer Science

TUCS Technical Reports

No 417, August, 2001

On Strongly Identifying Codes

Iiro Honkala

Department of Mathematics
University of Turku
FIN-20014 Turku, Finland
The research of this author was supported by the Academy of Finland under grant #44002.

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

Turku Centre for Computer Science
TUCS Technical Report No 417
August, 2001

ISBN 952-12-0865-1
ISSN 1239-1891

Abstract

Identifying codes are designed for locating faulty processors in multiprocessor systems. In this paper we consider a natural extension of this problem and introduce strongly identifying codes. Several lower bounds and constructions are given and relations between different types of identifying codes are examined.

Keywords: Identifying code, binary code, Hamming space, covering code

TUCS Research Group
Coding Theory Group

1 Introduction

Consider the Cartesian product $F_2^n = F_2 \times \cdots \times F_2$ of n copies of the binary field F_2. We endow this vector space with the Hamming metric; the Hamming distance $d(x, y)$ between vectors x and y is the number of coordinates in which they differ. The Hamming weight $w(x)$ of x is defined as $d(x, 0)$. We call the set $\{i \mid x_i \neq 0\}$ the *support* of $x = (x_1, \ldots, x_n)$. As usual, we denote by $|X|$ the cardinality of a set X,

$$B_t(x) = \{y \in F_2^n \mid d(x, y) \leq t\}$$

and

$$S_t(x) = \{y \in F_2^n \mid d(x, y) = t\}.$$

The following problem was introduced by Karpovsky, Chakrabarty and Levitin [15]. Let 2^n processors be arranged in the nodes of F_2^n. A processor can check all the processors within Hamming distance t and reports NO if something is wrong in these processors and YES otherwise. We want to choose a subset of processors (i.e. a subset of F_2^n) such that if there are problems in at most l processors, we know the locations of these malfunctioning processors by looking at the reports from the processors in the chosen subset.

In this model we expect to get correct reports from all the processors of the chosen subset. In other words, the faulty processors must be able to transmit the correct reports of their state as well. In this paper we consider the situation where malfunctioning processors send a report which may be correct or wrong. We can also think that only the processors with the "NO" answers transmit the report and a malfunctioning processor may send a report or be silent.

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

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

In order to find the malfunctioning processors we require that C satisfies the following. Let for any different subsets X and Y of F_2^n ($|X|, |Y| \leq l$) the sets $I_t(X) \setminus S$ and $I_t(Y) \setminus T$, where $S \subseteq X \cap C$ and $T \subseteq Y \cap C$, be nonempty and distinct. Then obviously 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. This leads to the following definition.

Definition 1. *Let $C \subseteq F_2^n$ be a code and $t, l \geq 0$ integers. For $X \subseteq F_2^n$ we define*

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

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 the definition of a (regular) $(t, \leq l)$-identifying code in the sense of Karpovsky, Chakrabarty and Levitin [15];

Definition 2. *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 $I_t(X_1) \neq I_t(X_2)$.*

Therefore, a strongly identifying code is always a (regular) identifying code. A strongly $(t, \leq l)$-identifying code is abbreviated by a *SID* code when the parameters t and l are known from the context, and we denote a strongly $(t, \leq 1)$-identifying code by a strongly t-identifying code.

We denote $I_t(\{x_1, \ldots, x_s\}) = I_t(x_1, \ldots, x_s)$ and $I'_t(y) = I_t(y) \setminus \{y\}$. In this paper we consider strongly $(t, \leq 1)$-identifying codes (the more general case $l \geq 2$ is examined in [16]). To verify that a code $C \subseteq F_2^n$ is $(t, \leq 1)$-identifying one must check that $I_t(y) \neq I_t(x) \neq I'_t(y)$ and $I'_t(x) \neq I'_t(y)$ for all $x, y \in F_2^n$ $(x \neq y)$ and that the sets $I_t(x)$ and $I'_t(x)$ are nonempty for all $x \in F_2^n$.

The smallest cardinality of a strongly t-identifying code of length n is denoted by $M_t^{SID}(n)$. Usually we omit t from these notations if $t = 1$. A code attaining the smallest cardinality is called *optimal*. We say that x t-covers y, if $d(x, y) \leq t$, and again we omit t, if $t = 1$.

2 Constructions

In what follows we often use the fact that three Hamming spheres of radius one intersect in a unique point, if the intersection is nonempty. Indeed, if the intersection contains a point, say x, then either x is one of the centres of the spheres or not. In both cases one immediately checks that the intersection contains only x.

The *direct sum* of the codes $C_1 \subseteq F_2^{n_1}$ and $C_2 \subseteq F_2^{n_2}$ is

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

Theorem 1. *For $n \geq 4$: $M^{SID}(n) \leq 2^{n-1}$.*

Proof. The set $F_2^{n-1} \oplus \{0\}$ is a strongly 1-identifying code. Here each codeword x is covered by exactly n codewords and every non-codeword $y = y'1$, where $y' \in F_2^{n-1}$ is covered by the unique codeword $y'0$. Thus clearly $I(y) \neq I(x) \neq I'(y)$ and $I'(x) \neq I'(y)$ for any distinct words x and y. □

Theorem 2. *Let $C \subseteq F_2^n$ be a $(1, \leq 1)$-identifying code with the property that $d(c, C \setminus \{c\}) = 1$ for all $c \in C$. Then $D = C \oplus F_2$ is strongly 1-identifying.*

Proof. We know by [1] that D is $(1, \leq 1)$-identifying. Let us now compare $I'(c)$, $c \in D$, to $I(x)$ and $I'(c')$ for all $x \in F_2^{n+1}$ and $c' \in D$ $(c \neq c')$. If c, x and c' are in $F_2^n \oplus \{1\}$, then $I(x) \neq I'(c) \neq I'(c')$ because of the unique word in $I'(c) \cap (F_2^n \oplus \{0\})$. Assume then that x and c' are in $F_2^n \oplus \{1\}$ and c is in $F_2^n \oplus \{0\}$. We have $I(x) \neq I'(c)$, since if $x \notin D$, then there exists $c_1 \in I'(c)$ in $F_2^n \oplus \{0\}$ such that $c_1 \notin I(x)$, and if on the other hand $x \in D$, then there exists $c_1 \in I(x)$ in $F_2^n \oplus \{1\}$ such that $c_1 \notin I'(c)$. Finally, $I'(c) \neq I'(c')$ because the contrary would only be possible if in C the corresponding words of F_2^n, say, c_C and c'_C were only covered by one another. However, since C is a $(1, \leq 1)$-identifying code, this cannot be true. □

Corollary 1. *If C is a strongly 1-identifying code, then its direct sum with F_2 is as well.*

Codes that are $(1, \leq 2)$-identifying have been considered in [14] and [18]. The smallest cardinality of a $(1, \leq 2)$-identifying code of length n is denoted by $M^{(\leq 2)}(n)$. In the next theorem we show that a $(1, \leq 2)$-identifying code is strongly $(1, \leq 1)$-identifying.

Theorem 3. $M^{SID}(n) \leq M^{(\leq 2)}(n)$.

Proof. Let C be a $(1, \leq 2)$-identifying code with $M^{(\leq 2)}(n)$ codewords. By [14, Theorem 2] every word in F_2^n is covered by at least three codewords. Thus for all $x \in F_2^n$ the set $I(x)$ is unique. Moreover, $I'(c) \neq I'(c')$ for all $c, c' \in C$ $(c \neq c')$. In particular, if $I'(c) = \{c_1, c_2\} = I'(c')$, then $I(c_1, c) = I(c_1, c')$ and this is a contradiction. □

Example 1. Let C be a code such that every word of the ambient space is covered by at least three codewords of C. As the previous proof indicates, C is then strongly 1-identifying provided that there does not exist a pair of codewords c and c' such that $I'(c) = \{c_1, c_2\} = I'(c')$. It can be checked (by computer) that this is the case with the code of length 13 and cardinality 1920 from [17], which gives the smallest currently known cardinality among the codes of length 13 such that every point in F_2^{13} is covered at least three times. Hence $M^{SID}(13) \leq 1920$.

The covering radius of a code C is defined via $R = \max_{x \in F_2^n} \min_{c \in C} d(x, c)$. Denote by $K(n, R)$ the smallest cardinality of a binary code of length n and covering radius R. Bounds for $K(n, R)$ can be found from [5].

Theorem 4. $M^{SID}(n) \leq 2n \cdot K(n-1, 2)$.

Proof. Let D be a code of length $n-1$ with covering radius two attaining $K(n-1, 2)$. Now $C' = F_2 \oplus D$ has the property that every word in F_2^n is at distance 0 or 2 from a codeword (this is not necessarily the closest codeword). Let $C' = \{c_1, c_2, \ldots, c_{|C'|}\}$.

We show that $C = \{a + e \mid a \in C', e \in S_1(0)\}$ is a SID code. Let $x \in F_2^n$ and denote $H_j = \{c_j + e \mid e \in S_1(0)\}$, $j = 1, \ldots, |C'|$. We have $d(x, c_j) = 0$ or 2 if and only if $|I(x) \cap H_j| \geq 2$ (and if and only if $|I'(x) \cap H_j| \geq 2$). Thus using $I(x)$ (or $I'(x)$) we can find a codeword c_j at distance zero or two (without loss of generality, we may assume $c_j = 0$). If $|I(x) \cap H_j| = n$, then $x = 0$. If $|I(x) \cap H_j| = 2$, then the union of supports of these two words is the support of x and again x is uniquely identified. □

The result above can also be proved using Theorem 2 and a modification of the code used to obtain [15, Theorem 8].

In the previous proof we constructed a code C' with the property

$$\bigcup_{c \in C'} (S_0(c) \cup S_2(0)) = F_2^n. \tag{2}$$

Can we construct a code satisfying (2), whose cardinality is smaller than $2K(n-1, 2)$? Following the techniques of [3] we next show that it is impossible. Let C be a code satisfying (2). Evidently, the requirement (2) is equivalent to the condition that every even-weight word is at distance zero or two from a codeword of even weight and every odd-weight word is at distance zero or two from an odd-weight codeword. Denote $C_e = \{c \in C \mid w(c) \text{ is even}\}$. Since the minimum distance of C_e is at least two, puncturing yields a code of length $n-1$, cardinality $|C_e|$ and covering radius at most two. This implies that $|C_e| \geq K(n-1, 2)$. Similarly, $|C \setminus C_e| \geq K(n-1, 2)$. Combining these we obtain $|C| \geq 2K(n-1, 2)$.

Example 2. Any $n-1$ codewords of weight one are enough to identify all words of weight two. By using this idea in the previous theorem when $n = 6$ we notice that we can remove four words. For example if $C' = \{000000, 100000, 011111, 111111\}$, then $C = \{a + e \mid a \in C', e \in S_1(0)\} \setminus \{110000, 100000, 001111, 011111\}$ is a SID code of cardinality 20. Here we have to notice that we cannot remove codewords freely; both 100000 and 000000 cannot be removed.

Theorem 5. *Assume that $2 \leq t < n/2$. Let $C' \subseteq F_2^n$ be a code with covering radius $2t$ attaining $K(n, 2t)$. Let further $B = B_{t-1}(0)$ and A consist of all the words of length n and weight t such that at least $t-1$ of the coordinate positions belong to the same residue class modulo two. Thus for every $2t$-element subset S of the set $\{1, 2, \ldots, n\}$ there is a collection of words of A such that the union of their supports is S. Then the code $C = C' + (A \cup B) = \{c + d \mid c \in C', d \in A \cup B\}$ is strongly t-identifying.*

Proof. Let us assume that x is the unknown word, and we are given a set $J(x)$, which is either $I(x)$ or $I'(x)$ (but of course we do not know which).

By the definition of C', there is a word $c \in C'$ such that $d(x, c) \leq 2t$. We easily find such a word c simply by checking whether or not $J(x) \cap (c + (A \cup B)) \neq \emptyset$. In particular, it is clear that all the sets $I(x)$ and $I'(x)$ are nonempty. Without loss of generality assume that $c = 0 \in C$ and that $d(x, c) \leq 2t$, i.e., $w(x) \leq 2t$.

Consider the smallest weight s occurring in $J(x)$. Clearly, $s = w(x) - t$ if $t < w(x) \leq 2t$; and $s = 0$ if $0 \leq w(x) \leq t$ — except when $x = 0$ and $J(x) = I'(x)$, but in this case $J(x)$ contains all the words of weight one, which is not true if $w(x) = t + 1$ (because $n \geq t + 2$), so this case can be disposed of. So, if $s > 0$, we can deduce that $w(x) = s + t$, and obtain the support of x as the union of the supports of the codewords of weight s in $J(x)$. Assume therefore that $s = 0$ (i.e., $0 \in J(x)$). Then $0 \leq w(x) \leq t$.

Let now ℓ be the largest weight such that at most one of the words of $A \cup B$ of that weight is missing from $J(x)$. Because $n \geq 2t + 1$ (and hence $n \geq t + 2$), we see that $\ell = t - w(x)$. The case $w(x) = 0$ is now clear. Let us first assume that $w(x) \neq (t + 1)/2$. Then we know that $w(x) \neq \ell + 1$, and (even in the case $J(x) = I'(x)$) the union of supports of the words of weight $\ell + 1$ that are not in $J(x)$ is exactly the complement of the support of x. Indeed, if $2 \leq w(x) \leq t$, then this is immediate; if $w(x) = 1$, we use the fact that every $2t$-element subset with an empty intersection with the support of x can be obtained as a union of supports of some words in A. Assume finally that $w(x) = \ell + 1 = (t + 1)/2$. We know that $1 < (t + 1)/2 < t$, and $n \geq t + 2$. If i is an element which is not in the support of x, at least two of the words of weight $\ell + 1$ that are not in $J(x)$ contain i; if i is in the support of x, there is at most one. Again we find the support of x. □

The fact that the code C in the previous theorem is (regular) $(t, \leq 1)$-identifying was already shown in [2].

No asymptotically better bounds than the ones in Theorems 4 and 5 are known even for (regular) identifying codes [15, 2].

Definition 3. *Denote by $W_t(n,k,l)$ the minimum number of codewords in any code C of length n whose codewords all have weight l and which has the property that all the sets $I_t(C;x)$, $x \in S_k(0)$, are nonempty and different.*

Trivially $W_1(n,2,1) = n-1$. Let us look at the values of $W_1(n,3,2)$.

If we denote by $D(C)$ the smallest number of different codewords of C whose sum is the all-zero word, we get the following theorem.

Theorem 6. *Let $\bar{C} = S_2(0) \setminus C$, $C \subseteq S_2(0)$. The sets $I(\bar{C},x)$ are nonempty and distinct for all words x of weight three if and only if $D(C) \geq 5$.*

Proof. Clearly, there is a word x of weight three for which $I(\bar{C},x) = \emptyset$ if and only if C contains three words whose sum is the all-zero word. Any word $x \in S_3(0)$ for which $|I(\bar{C},x)| \geq 2$ is obviously uniquely identified. We only need to check whether there exist cases in which for two words of weight three, say x_1 and x_2 with supports $\{i,j,k\}$ and $\{i,j,m\}$, we have $I(\bar{C},x_1) = \{c\} = I(\bar{C},x_2)$ where the support of c equals $\{i,j\}$. This happens if and only if the words with the supports $\{i,k\}$, $\{j,k\}$, $\{i,m\}$ and $\{j,m\}$ are all in C, i.e., if and only if some four codewords of C add up to the all-zero word. $\square$

According to the theorem above, determining the values of $W_1(n,3,2)$ is equivalent to finding the largest code $C \subseteq F_2^n$ whose codewords are all of weight two and $D(C) \geq 5$. Consider an undirected graph whose vertex set is the set of the coordinates $\{1,\ldots n\}$ and an edge is a pair of such coordinates (that is, the support of a word of weight two). Hence calculating $W_1(n,3,2)$ is equivalent to finding a graph with the maximal number of edges and with the length of the shortest cycle (girth) at least five.

The problem of finding such graphs is well-known and several exact values of $W_1(n,3,2)$ are known (see, e.g., [11, 12, 19] and the references therein). For example, the values of $W_1(n,3,2)$ are 1, 3, 5, 9, 13, 18, 24, 30, 39, 48, 57, 68, 79, 92, 105, 119 for the lengths $n = 3,\ldots 18$, respectively. It also follows that

$$\lim_{n\to\infty} \frac{W_1(n,3,2)}{n^2} = \frac{1}{2}.$$

Example 3. The constant weight code $S_2(0)\setminus$ {1100000, 0110000, 0011000, 0001100, 0000110, 1000010, 1000001, 0001001} attains the value $W_1(7,3,2) = 13$.

Theorem 7. *Suppose $n \geq 7$. If A is a code attaining the value $W_1(n,3,2)$, then every word of weight one is covered by at least three codewords of A.*

Proof. Suppose on the contrary that there is a word x of weight one, which is covered by less than three codewords. Let $\{s\}$ be the support of x. Without loss of generality we can assume that $s = 1$. Denote by i the number of codewords of weight two that cover x.

If $i = 0$, then none of the words with support $\{1, j\}$, for $j = 2, \ldots, n$, is a codeword. There are $n - 1$ such words. One easily checks that the code consisting of all the words of length $n \geq 8$ and weight two except the ones with supports $\{1, 2\}$, $\{2, 3\}, \ldots, \{n-1, n\}$, $\{n, 1\}$ and $\{1, 5\}$ still identifies all the words of weight three. From this and Example 3 we can deduce that for $n \geq 7$ there are at least $n + 1$ words of weight two that do not belong to A. Hence there is also a word whose support is $\{k, l\}$, $(k, l \geq 2,\ k \neq l)$, which is not in A. But now the word of weight three with support $\{1, k, l\}$ is not covered at all, a contradiction.

If $i = 1$, then for some j the word with support $\{1, j\}$ is a codeword. Now there must be also three words of weight two which begin with zero and are not codewords. If any of them does not contain j in its support then we are done as in the previous case. Without loss of generality words with supports $\{j, k_1\}$ and $\{j, k_2\}$ are not codewords, for some k_1 and k_2, $k_1 \neq k_2$. But now the words of weight three with supports $\{1, j, k_1\}$ and $\{1, j, k_2\}$ cannot be distinguished, because both are only covered by the word with support $\{1, j\}$.

If $i = 2$, then for some j_1 and j_2 words with supports $\{1, j_1\}$ and $\{1, j_2\}$ are codewords. Now there are at least four words of weight two that begin with zero and are not codewords. As in the previous case either j_1 or j_2 must occur in the support of each of them. This means without loss of generality that there are words with supports $\{j_1, k_1\}$ and $\{j_1, k_2\}$, $k_1 \neq j_2$, $k_2 \neq j_2$, $k_1 \neq k_2$, that are not codewords. (The remaining two noncodewords can have supports $\{j_1, j_2\}$ and $\{j_2, k\}$.) And we get a contradiction as in the previous case. ☐

Theorem 8. *For $n \geq 7$,*

$$M^{SID}(n) \leq (W_1(n, 3, 2) + n - 1)K(n, 3).$$

Proof. Let A be a set which attains the value $W_1(n, 3, 2)$, B realizing the value $W_1(n, 2, 1) = n - 1$, and D a code of length n and covering radius three. We show that $(A \cup B) + D$ is SID.

Let $x \in F_2^n$ and denote $H(d) = \{d + y \mid y \in A \cup B\}$. Then $d(x, d) \leq 3$ if and only if $H(d) \cap I(x) \neq \emptyset$ (and if and only if $H(d) \cap I'(x) \neq \emptyset$). Therefore, using $I(x)$ (or $I'(x)$) we can find a codeword $d \in D$ such that $d(x, d) \leq 3$. Without loss of generality, assume that $d = 0$.

If $x = 0$, then we immediately know it, because $I(x)$ and $I'(x)$ both contain at least $n-1 \geq 3$ words of weight one, which uniquely identify x. Assume that $x \neq 0$. If $I(x)$ (or $I'(x)$) contains 0, then we know that $w(x) = 1$, and by Theorem 7, at least three of the words in A cover x, and therefore uniquely identify it. We can now assume that we know that $w(x) \geq 2$. Then $w(x) = 2$ if and only if $I(x)$ (or $I'(x)$) contains at least one word of weight one. When it is known that $w(x) = 2$, the words in B uniquely identify x. When we know that $w(x) = 3$, then words of A uniquely identify x. □

3 Nonexistence results

Denote by N_i the number of codewords of weight i in C.

Lemma 1. *Let C be a strongly 1-identifying code of length $n \geq 3$. If $0 \notin C$ then at least $\lceil 2n/3 \rceil$ codewords of weight two are needed to identify all the words of weight one. If $0 \in C$, then we need at least $\lceil 2(n-1)/3 \rceil$ codewords of weight two.*

Proof. Assume first that $0 \notin C$. If s denotes the number of words $x \in S_1(0)$ such that $|I'(x)| = 1$, then

$$s + 2(n-s) \leq \sum_{x \in S_1(0)} |I'(x)| = 2N_2.$$

Since $s \leq N_2$, we get $N_2 \geq \lceil 2n/3 \rceil$.

Assume then that $0 \in C$. Then $I'(x) = \{0\}$ for at most one $x \in S_1(0)$. Considering the sets $I'(x) \setminus \{0\}$, we similarly get

$$N_2 + 2(n - N_2 - 1) \leq \sum_{x \in S_1(0)} |I'(x) \setminus \{0\}| = 2N_2,$$

and the second claim follows. □

Theorem 9.

$$M^{SID}(n) \geq \left\lceil \frac{2^n \cdot \lceil 2n/3 \rceil}{\binom{n}{2} + \lceil 2n/3 \rceil - \lceil 2(n-1)/3 \rceil} \right\rceil.$$

Proof. Assume that C is a code with $M^{SID}(n)$ codewords. Applying Lemma 1 to all the words of F_2^n we get

$$(2^n - M^{SID}(n))\lceil 2n/3 \rceil + M^{SID}(n)\lceil 2(n-1)/3 \rceil \leq M^{SID}(n)\binom{n}{2},$$

from which the theorem follows. □

Lemma 2. *If C is an optimal strongly 1-identifying code of length at least four, then for all $x \in F_2^n$ we have $|I'(x)| \leq n-1$.*

Proof. Suppose on the contrary that for some x we have $|I'(C;x)| = n$. Without loss of generality we can assume that $x = 0$. We will show that $C \setminus \{y\}$ is a SID code, we choose y in the following way: we take $y \in S_1(0)$ such that $|I(C;y) \cap S_2(0)| = 1$, if such a word exists; otherwise we take any $y \in S_1(0)$. It suffices to show that

$$I'(v) \neq I(w) \neq I(v) \text{ and } I'(v) \neq I'(w) \neq I(v) \tag{3}$$

for all $w \in B_1(y)$ and $v \in F_2^n$. Here and from now on the notations I and I' all refer to the code $C \setminus \{y\}$. Since $|I'(0)| = n-1 \geq 3$ we may always exclude the cases $w = 0$ and $v = 0$.

Assume first that also $v \in B_1(y)$ and $0 \neq v \neq w \neq 0$. Then either $w \neq y$ or $v \neq y$; say $w \neq y$. Now $w + y \in I'(w) \cap S_1(0)$ but $w + y \notin I(v)$. This implies (3) in this case.

Suppose then that $v \notin B_1(y)$. Let first $w(v) \geq 3$. Since C is SID, we get (3) for $w = y$. If $w \neq y$, $I'(w) \cap S_1(0) \neq \emptyset$ and $I(v) \cap S_1(0) = \emptyset$. This gives (3) for these v. Let then $w(v) = 2$. Now there clearly exists a codeword $a \in I'(v) \cap S_1(0)$ such that $a \notin I(w)$ (because $w \neq 0$). Let finally $w(v) = 1$. Since $I'(C;v) \neq I'(C;y)$ we may assume that $w(w) = 2$. The choice of y guarantees that there exists a word of weight two in $I'(v)$ which does not belong to $I(w)$. Therefore, we have (3) for all w and v. □

In the proof of the next lower bound (cf. [15, Theorem 3] and [1, Theorem 9]) we use the concept of *excess*, cf., e.g., [5]: Assume that $C \subseteq F_2^n$ has covering radius one. If a vector $x \in F_2^n$ is 1-covered by exactly $i+1$ codewords of C then we say that the excess $E(x)$ on x is i. In general, the excess $E(V)$ on a subset $V \subseteq F_2^n$ is defined by $E(V) = \sum_{x \in V} E(x)$.

Theorem 10.

$$M^{SID}(6) \geq 18,\ M^{SID}(7) \geq 32,\ M^{SID}(8) \geq 57.$$

For $n \geq 9$,

$$M^{SID}(n) \geq \left\lceil \frac{2^{n+1}(n^2 - 2n + 4)}{n^3 - n^2 + 2n + 8} \right\rceil.$$

Proof. Let C be a code realizing $M^{SID}(n)$ with $n \geq 6$. The number of points for which $|I(x)| = 1$ is at most $M^{SID}(n)$. The points x with $|I(x)| = 2$ are called *sons* and the ones with $|I(x)| > 2$ are called *fathers*. If $|I(x)| = 2$, then there exists a unique point y such that $I(x) \subset I(y)$ and is called the father of x. A *family* consists of a father and its sons. The space is partitioned by the families and the points with $|I(x)| = 1$.

Assume that f is a father and denote by $S = S(f)$ the set of sons of f. Let $|I(f)| = i \geq 3$. We shall examine the average excess of a family, i.e., the function

$$\frac{|S| + i - 1}{|S| + 1} =: g(i, |S|).$$

Let us bound above the number of sons of a father and thus bound below the average excess of a family. Without loss of generality we can assume that the father f is the all-zero word. Clearly, a father may have at most $\binom{i}{2}$ sons, but as we shall see, we can often say more.

Assume first that $f \notin C$. By the previous lemma we know that $i \leq n-1$. Suppose first that $i = n - 1$.

Denote by x the unique word not in $I(f)$ but for which $d(f, x) = 1$. We show that all the $n - 2$ words in $B_1(c) \setminus \{f, c, x + c\}$ cannot be sons of f. Indeed, there must be a codeword $c' \neq c$ in $I(c)$, otherwise $I'(c) = \emptyset$, and if $c' \neq x + c$ we are done, so assume $c' = x + c$ is the unique codeword in $I'(c)$. There has to be a codeword c'' of weight three in $I(c')$: otherwise $I(c') = I(c)$. We have $c'' = c + x + z$ for some $z \in I(f)$. But then $c + z$ cannot be a son, since $|I(c + x)| \geq 3$. Consequently, $c \in I(f)$ can have at most $n - 3$ sons of f at distance one from it. Counting in two ways the pairs (c, s) where $c \in I(f)$, $s \in S$ and $d(c, s) = 1$ we obtain $2|S| \leq (n-3)|I(f)|$ which implies $|S| \leq \lfloor (n-3)(n-1)/2 \rfloor =: U_1$.

Consider the case $f \notin C$ and $i = n - 2$. Denote by x_1 and x_2 the two words not in $I(f)$ but at distance one from f. Because C is SID, $I'(x_1 + c) \neq I'(x_2 + c)$, and therefore one of these contains a codeword of C of weight three, which is not contained in the other. Without loss of generality, $x_1 + c + z \in C$ for some $z \in I(f)$, $z \neq c$. Then $c, z, x_1 + c + z \in I(c + z)$ and hence $c + z$ is not a son. Thus there cannot be $n - 3$ sons in $S_1(c)$. Counting as above, we get $|S| \leq \lfloor (n-2)(n-4)/2 \rfloor =: U_2$.

Notice that for other values $i = 4, \ldots n$ the function $g(i, \binom{i}{2})$ is decreasing, and $g(3, \binom{3}{2}) = g(6, \binom{6}{2})$. Hence for $i = 3, \ldots, n-3$ we may bound $g(i, \binom{i}{2})$ below by $g_1(n) := g(n-3, \binom{n-3}{2})$, when $n \geq 9$.

Assume now that $f \in C$ and $|I(f)| = i$. In this case S consists of the sons at distance one and two from f. Since a son is covered by exactly two codewords, there can be only one son at distance one from f. Indeed, if $s_1, s_2 \in S$ and $d(f, s_1) = d(f, s_2) = 1$, then $I'(s_1) = I'(s_2) = \{f\}$. Consequently, $|S| \leq \binom{i-1}{2} + 1$, because there are at most $\binom{i-1}{2}$ sons of weight two.

Notice that $g(i, \binom{i-1}{2} + 1)$ is decreasing on i ($i = 4, \ldots, n$) and we may bound it below by $g_2(n) := g(n, \binom{n-1}{2} + 1)$ for $n \geq 6$.

The minimum of $g(3, \binom{3}{2}), g(4, \binom{4}{2}), g(5, \binom{5}{2})$ and $g_2(3)$ is $g(3, \binom{3}{2}) = 5/4$. So to find the minimum of the above mentioned lower estimates on $g(i, |S|)$ we need to compare the functions $g_1(n), g(n-1, U_1), g(n-2, U_2), g_2(n)$, and $5/4$. When $6 \leq n \leq 8$ the minimum is $5/4$ and for $n \geq 9$ the minimum is $g_2(n)$. Denote by $M(n)$ the minimum.

Consequently, the average excess of a family is at least $M(n)$ and hence for every family $\mathcal{F}$ the excess of it $E(\mathcal{F}) \geq M(n)|\mathcal{F}|$. Since the excess on F_2^n by C is $(n+1)M^{SID}(n) - 2^n$ we get

$$(n+1)M^{SID}(n) - 2^n \geq (2^n - M^{SID}(n))M(n).$$

Routine calculations give the claim. □

There exist (see, [20, Construction 4.24]) infinite sequences of codes $(C_i)_{i=1}^{\infty}$ of length $n_i \to \infty$ and covering radius two such that

$$\lim_{i\to\infty} \frac{K(n_i, 2)|B_2(0)|}{2^{n_i}} = 1.$$

Such a family exists, for example, for the lengths $n_i = 2^i + 5 \cdot 2^{i/2-2} - 2$ where $i \geq 4$ is even.

Combining this result with Theorem 4 and Theorem 10, we have an infinite sequence of strongly identifying codes such that the ratio between their cardinalities and the lower bound of Theorem 10 approaches one. Using the results of [15], we also see that we have an infinite sequence of lengths such that asymptotically the ratio between the smallest cardinalities of identifying codes and strongly identifying codes tends to one.

4 Short codes

Theorem 11. $M^{SID}(3) = 6$, $M^{SID}(4) = 8$.

Proof. The lower bound on $M^{SID}(3)$ follows from Theorem 9 and the upper bound from Theorem 4.

The upper bound in the case $n = 4$ comes from Theorem 1. Suppose that $M^{SID}(4) \leq 7$ and let C be a code attaining the value $M^{SID}(4)$. Throughout the proof we will only be using the condition that the sets $I'(x)$ are nonempty and different (a fact that we will need in the last section). We can assume that neither 0000 nor 1111 is a codeword. Namely, if every word-complement pair contains at least codeword, there would be at least eight codewords.

By Lemma 1 we need at least $\lceil 2 \cdot 4/3 \rceil = 3$ codewords of weight two to identify all words of weight one. Assume that $N_2 = 3$. These codewords can be chosen in two ways. Either there are two words of weight one that are covered by two codewords, or there is a word which is covered by three codewords; in both cases all the other words of weight one are covered by one codeword each. In the first case if x is a word of weight one such that $|I'(x)| = 2$ then there exists a word y of weight three such that $I'(y) = I'(x)$, so we need at least four codewords of weight two. In the second case if x is the word for which $|I'(x)| = 3$ then for the complement $\overline{x}$ of x we have $I'(\overline{x}) = \emptyset$, so again $N_2 \geq 4$.

Because $I'(0000) \neq \emptyset$ and $I'(1111) \neq \emptyset$, we have $N_1 \geq 1$ and $N_3 \geq 1$. Because $I'(x) \neq \emptyset$ for all x of weight two, there is a codeword c of weight one such that its complement is also a codeword, or $N_1 + N_3 \geq 4$. Without loss of generality we can assume that 1000 and 0111 are codewords. Because $I'(1100)$, $I'(1010)$ and $I'(1001)$ must all be different, we need at least two more codewords of weight one or three. Again $N_1 + N_3 \geq 4$. □

Theorem 12. $M^{SID}(5) = 14$.

Proof. A strongly 1-identifying code of length 5 and cardinality 14 is {10000, 01000, 00100, 11000, 01100, 01010, 00011, 11010, 10101, 01101, 01011, 11101, 11011, 10111}.

Throughout the proof of the lower bound we will again only be using the condition that the sets $I'(x)$ are nonempty and different.

Let C be a strongly 1-identifying code of length five. To prove the lower bound we can assume that neither 00000 nor 11111 is a codeword. Namely, if every word-complement pair contained at least one codeword, then there would be at least 16 codewords. We will prove that $N_1 + N_3 \geq 7$. By symmetry (considering the code $\{1 + c | c \in C\}$) we then know that also $N_2 + N_4 \geq 7$, and the claim follows. Because $I'(00000)$ is not empty, $N_1 \geq 1$.

Case 1: $N_1 = 1$. We can assume that $10000 \in C$. Denote $A = \{11000, 10100, 10010, 10001\}$. We know the sets $I'(x)$, $x \in A$ are different and, because $I'(00000) = \{10000\}$, each contain at least one codeword of weight three. Hence C contains at least three codewords of weight three that begin with 1, say c_1, c_2 and c_3. Each of these three codewords covers one word of weight two that is not in A. But still we have three words of weight two that are not covered by them. If c_1, c_2 and c_3 are at distance one from one word in A, then the remaining three words of weight two that are not covered, say x_1, x_2 and x_3, are all at distance one from one word of weight three. Because $I'(x_1)$, $I'(x_2)$ and $I'(x_3)$ must be nonempty and different, there must be at least three more codewords of weight three, and hence $N_1 + N_3 \geq 7$. Suppose therefore that two words from the set A are covered by two of the codewords c_1, c_2 and c_3 and two by one. Without loss of generality we can assume that c_1, c_2 and c_3 are $11100, 10110$ and 10011. Now the words $01010, 01001$ and 00101 are not covered. The only possibility to make $I'(01010)$, $I'(01001)$ and $I'(00101)$ all nonempty and different using only two more codewords of weight three is to choose 01101 and 01011 as codewords. But then $I'(01111) = I'(01001)$, so we need all in all at least six codewords of weight three, and hence $N_1 + N_3 \geq 7$.

Case 2: $N_1 = 2$. Without loss of generality assume $10000, 01000 \in C$. Because $I'(00000) \neq I'(11000)$ we can assume $11100 \in C$. Because $I'(10010) \neq I'(10001)$, at least one of the words 11010, 11001, 10110 or 10101 must be in C. The first two and the last two cases are symmetric. In the first case $N_1 + N_3 \geq 7$, because $I'(00110)$, $I'(00101)$ and $I'(00011)$ are nonempty and different, and we need three more codewords of weight three. In the last case suppose that 10110 is a codeword. Again $I'(00110)$, $I'(00101)$ and $I'(00011)$ must be nonempty and different, which requires three codewords of weight three, but of course 10110 can be used as one of them. Let us assume that $N_3 = 5$. Then of the remaining two codewords of weight three (at least) one has to cover either 01010 or 01001 but not both. There are now three possibilities to choose them: 1) 00111, 01101, now $I'(01111) = I'(00101)$, 2) 01101, 01011, now $I'(11011) = I'(00011)$, 3) 01101, 10011, now $I'(01111) = I'(00101)$. So in each case we need at least one more codeword of weight three, and so $N_1 + N_3 \geq 7$.

Case 3: $N_1 \geq 3$. By Lemma 1 we know that $N_3 \geq 4$, so $N_1 + N_3 \geq 7$. □

For the usual $(1, \leq 1)$-identifying codes of lengths 2, 3, 4 and 5 the smallest cardinalities are 3, 4, 7 and 10, respectively. Other values of the cardinalities of $(1, \leq 1)$-identifying codes can be found from [1].

lower bounds	upper bounds
a Theorem 9	A Theorem 11
b Theorem 11	B Theorem 1
c Theorem 12	C Theorem 12
d Theorem 10	D Example 2
	E Theorem 8
	F Theorem 2
	G Theorem 3
	H Theorem 4
	I Example 1

n	$M^{SID}(n)$
3	a 6 A,F
4	b 8 B,F
5	c 14 C,F
6	d 18–20 D,F
7	d 32–38 E
8	d 57–64 F
9	d 102–128 F
10	d 186–256 F
11	d 341–512 G,F
12	d 629–1024 G,F
13	d 1169–1920 I
14	d 2182–3584 H
15	d 4091–6144 G

Table 1: Bounds on $M^{SID}(n)$.

5 On $(t, \leq l)$-identifying code with nontransmitting faulty vertices

In the introduction we considered faulty processors, which did or did not send the report on their neighbourhood. A malfunctioning processor may also *always* be unable to transmit the report and thus we always get incorrect information about the neighbourhood of this processor. In this section we briefly discuss this variant.

Definition 4. *Let $C, X \subseteq F_2^n$. Define $\mathcal{I}_t(X) = I_t(X) \setminus (X \cap C)$. If for all distinct subsets $X \subseteq F_2^n$ of cardinality at most l the sets $\mathcal{I}_t(X)$ are different, then the code C is called a $(t, \leq l)$-identifying code with nontransmitting faulty vertices.*

A strongly $(t, \leq l)$-identifying code is also a $(t, \leq l)$-identifying code with nontransmitting faulty vertices by the definitions. The smallest cardinality of a $(1, \leq 1)$-identifying code of length n with nontransmitting faulty vertices (abbreviated to an *IDNT code*) is denoted by $M^{IDNT}(n)$.

Theorem 13. *A $(1, \leq 1)$-identifying code with nontransmitting faulty vertices is a strongly $(1, \leq 1)$-identifying code if and only if there are no codewords c_1 and c_2 such that $I(c_1) = I(c_2) = \{c_1, c_2\}$.*

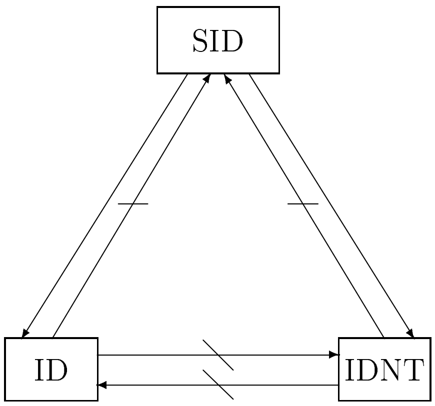

Figure 1: The relationships between the three concepts: cf. Example 4.

Proof. Let C be IDNT. If C is also SID, then clearly no such codewords c_1 and c_2 exist.

Let us verify the other direction. Since $I'(x) \neq I'(y)$ by the definition of IDNT, it suffices to check that $I(x) \neq I(y)$ and $I(x) \neq I'(y)$ for all distinct words x and y.

Assume that $I(x) = I'(y)$ for some x and y. We may assume that $x \in C$, otherwise we are done by the IDNT property. Now $I'(y) = I'(x) \cup \{x\}$ and thus $d(x, y) = 1$ and $B_1(x) \cap B_1(y) = \{x, y\}$. Since $I'(x) \neq \emptyset$, we get a contradiction.

Suppose then that $I(x) = I(y)$. Now we may assume that $x, y \in C$, otherwise we are done by the arguments above. Again $d(x, y) = 1$ and we get a contradiction by the fact that there are no codewords x and y such that $I(x) = I(y) = \{x, y\}$. □

The direct sum of an IDNT code C and F_2 leads to a new IDNT code if and only if C is a SID code. Namely, if in C there are codewords c_1 and c_2 such that $I(c_1) = \{c_1, c_2\} = I(c_2)$ then in $C \oplus \{0, 1\}$ there are four codewords $c_1 0, c_1 1, c_2 0$ and $c_2 1$ for which $I'(c_1 0) = I'(c_2 1) = \{c_1 1, c_2 0\}$, and thus $C \oplus \{0, 1\}$ is not an IDNT code.

Example 4. Let $C_1 = \{000, 110, 101, 011\}$ and

$$C_2 = \{0, 1\} \oplus \{x \in F_2^5 | w(x) \neq 1\}.$$

The code C_2 is IDNT, but not $(1, \leq 1)$-identifying, because the words 000000 and 100000 cannot be distinguished. Of course, C_2 is not strongly $(1, \leq 1)$-identifying either. A $(1, \leq 1)$-identifying code is not always IDNT either, which is clear from C_1. This code also shows that a $(1, \leq 1)$-identifying code is not always SID. The code C_2 also illustrates that $C \oplus \{0,1\}$ may be IDNT, although C is not.

Theorem 14. $M^{IDNT}(3) = 6,\ M^{IDNT}(4) = 8,\ M^{IDNT}(5) = 14.$

Proof. Lemma 1 holds for IDNT codes, thus the lower bound of Theorem 9 is still valid. Hence $M^{IDNT}(3) \geq 6$. For lengths four and five we have seen that the proofs of the lower bounds in Theorems 11 and 12 also work in the IDNT case. The theorem now follows because strongly identifying codes are IDNT codes. □

References

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

[2] U. Blass, I. Honkala, S. Litsyn, On binary codes for identification, Journal of Combinatorial Designs 8 (2000) 151–156.

[3] G. Cohen, P. Frankl, On tilings of the binary vector space, Discrete Math. 31 (1980) 271–277.

[4] G. Cohen, S. Gravier, I. Honkala, A. Lobstein, M. Mollard, C. Payan, G. Zémor, Improved identifying codes for the grid, Electronic Journal of Combinatorics, Comments to 6(1), R19, (1999).

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

[6] G. Cohen, I. Honkala, A. Lobstein, G. Zémor, New bounds for codes identifying vertices in graphs, Electronic Journal of Combinatorics 6(1), R19, (1999).

[7] G. Cohen, I. Honkala, A. Lobstein, G. Zémor, On codes identifying vertices in the two-dimensional square lattice with diagonals, IEEE Transactions on Computers 50 (2001) 174–176.

[8] G. Cohen, I. Honkala, A. Lobstein, G. Zémor, Bounds for codes identifying vertices in the hexagonal grid, SIAM J. Discr. Math., 13 (2000) 492–504.

[9] G. Cohen, I. Honkala, A. Lobstein, 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), pp. 97–109.

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

[11] D. K. Garnick, Y. H. H. Kwong, F. Lazebnik, Extremal graphs without three-cycles or four-cycles, J. Graph Theory 17 (1993) 633–645.

[12] D. Garnick, N. A. Nieuwejaar, Nonisomorphic extremal graphs without three-cycles or four-cycles, J. Combin. Math. Combin. Comput. 12 (1992) 33–56.

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

[14] I. Honkala, T. Laihonen, S. Ranto, On codes identifying sets of vertices in Hamming spaces, Designs, Codes and Cryptography, to appear.

[15] 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.

[16] T. Laihonen, S. Ranto, Codes identifying sets of vertices, Lecture Notes in Computer Science, Proceedings of AAECC-14, to appear.

[17] P. R. J. Östergård: New multiple covering codes by tabu search, Australasian J. Combinatorics, 12 (1995) 145–155.

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

[19] N. J. A. Sloane, The On-Line Encyclopedia of Integer Sequences, Published electronically at `http://www.research.att.com/~njas/sequences`.

[20] R. Struik, Covering Codes, Ph. D. Thesis, Eindhoven University of Technology, the Netherlands, 106 pp., 1994.

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**