

## ON COMPUTING MAJORITY BY COMPARISONS

MICHAEL E. SAKS<sup>1</sup> and MICHAEL WERMAN

*Received January 20, 1989*

*Revised February 6, 1990*

The elements of a finite set  $X$  (of odd cardinality  $n$ ) are divided into two (as yet unknown) classes and a member of the larger class is to be identified. The basic operation is to test whether two objects are in the same class. We show that  $n - B(n)$  comparisons are necessary and sufficient in worst case, where  $B(n)$  is the number of 1's in the binary expansion of  $n$ .

### 1. Introduction

Let  $X$  be a finite set of cardinality  $n$  ( $n$  odd) whose elements are partitioned in an unknown way into two parts  $A$  and  $B$ . The problem is to identify at least one element that belongs to the larger part. The basic operation permitted is a comparison of any two elements  $x, y \in X$ . The comparison is denoted  $x : y$  and its outcome is either " $x = y$ " or " $x \neq y$ ". Since comparing every element to a fixed element  $a \in X$  completely determines the partition,  $n - 1$  comparisons suffice to solve the problem. There is a comparison strategy that solves the problem using at most  $n - B(n)$  comparisons, where  $B(n)$  is the number of ones in the binary expansion of  $n$  (Such a strategy is described in section 3). The main result of this note is that this is tight.

**Theorem 1.1.**  $n - B(n)$  comparisons are necessary and sufficient to identify a member of the majority class.

The upper bound has apparently been observed by several people. The new result of this paper is the lower bound. It is proved by a generating function argument related to that introduced by Best, van Emde Boas and Lenstra ([1]) and Rivest and Vuillemin ([5]) (see [2], Chap. 8, Theorem 2.1) for analyzing the number of variables that must be probed to evaluate a Boolean function.

The problem was presented by Steve Grantham at the AMS-INS-SIAM meeting on "Graphs and Algorithms" June, 1987, but the original source is unclear. In the more general situation in which the variables take on arbitrary values, Fisher and Salzberg ([3]) showed that  $\lceil (3n/2) + 1 \rceil$  comparisons are necessary and sufficient to

---

AMS subject classification (1991): 68 P 10, 68 R 05

<sup>1</sup> Supported in part by NSF grant DMS87 03541 and Air Force Office of Scientific Research grant AFOSR-0271.

identify a majority element if one exists. Also, for general values of the variables and arbitrary  $k$ , Misra and Gries ([4]) obtained estimates of the number of comparisons needed to determine whether there are at least  $k$  variables with the same value.

## 2. Reformulation of the problem

Consider the state of knowledge after some set  $C$  of comparisons has been performed. The set  $C$  of comparisons can be divided into two sets  $C_ =$  and  $C_ \neq$  according to the results. Let  $\chi(C_ =, C_ \neq)$  be the set of partitions  $\{A', B'\}$  of  $X$  that are consistent with this information.

The condition for terminating the algorithm is:

(T) There exists an element  $x$  such that for every partition in  $\chi(C_ =, C_ \neq)$ ,  $x$  belongs to the larger part.

Let  $X_1, \dots, X_k$  be the partition of  $X$  into the connected components of the graph with edge set  $C$ . For each  $i \in [k] (= \{1, \dots, k\})$  the results of the comparisons within  $X_i$  determine the restriction  $\{A \cap X_i, B \cap X_i\}$  of the unknown partition  $\{A, B\}$  to  $X_i$ . Let  $(L_i, S_i)$  be the partition of  $X_i$  ordered so that  $|L_i| \geq |S_i|$ , and let  $m_i = |L_i| - |S_i|$ . For  $I \subseteq [k]$ , let  $L_I = \bigcup_{i \in I} L_i$ ,  $S_I = \bigcup_{i \in I} S_i$  and  $m_I = \sum_{i \in I} m_i$ . Then it is easily seen that

$$\chi(C_ =, C_ \neq) = \{\{L_I \cup S_{[k]-I}, L_{[k]-I} \cup S_I\} \mid I \subseteq [k]\}.$$

Furthermore,  $|L_I \cup S_{[k]-I}| - |L_{[k]-I} \cup S_I| = m_I - m_{[k]-I}$ . Thus an element  $x$  satisfies (T) if and only if  $x$  belongs to  $L_j$  for some  $j$  and  $m_I \geq m_{[k]-I}$  for all  $I$  containing  $j$ . From this we obtain the following equivalent termination condition:

(T') There exists  $j$  such that  $m_j \geq \sum_{i \neq j} m_i$ .

Note that further comparisons between two elements in the same  $X_i$  provide no new information. A comparison between any  $x \in X_i$  and  $y \in X_j$  determines  $\{A \cap (X_i \cup X_j), B \cap (X_i \cup X_j)\}$  to be either  $\{L_i \cup L_j, S_i \cup S_j\}$  or  $\{L_i \cup S_j, L_j \cup S_i\}$ , and hence causes  $m_i$  and  $m_j$  to be replaced by either  $m_i + m_j$  or  $|m_i - m_j|$ .

Thus the testing problem can be reformulated as a game between two players, the *selector* and the *assigner*. A position of the game is a multiset  $M = \{m_1, m_2, \dots, m_k\}$  of nonnegative integers. In each round of the game, the *selector* selects two members of the multiset and the *assigner* replaces these two numbers by either their sum or by their absolute difference, reducing the size of the multiset by 1. The game ends when the largest number in the multiset is at least half the total. The selector seeks to maximize (and the assigner to minimize) the size of the multiset when the game ends. The value of the game  $V(m_1, \dots, m_k)$  is the size of the set remaining if both players follow their optimal strategy. For  $n$  odd, let  $V_n = V(\underbrace{1, 1, \dots, 1}_{n \text{ times}})$ . Theorem 1.1. is then equivalent to:

**Theorem 2.1.** For  $n$  odd,  $V_n = B(n)$ .

In the next section we present a strategy for the selector which will prove  $V_n \geq B(n)$ . The main result, that  $V_n \leq B(n)$  is proved in section 4.