On Sorting by Comparisons

R. L. GRAHAM

*Bell Telephone Laboratories, Incorporated,
Murray Hill, New Jersey, U.S.A.*

1. Introduction

A problem which frequently occurs in the study of sorting algorithms can be phrased as follows. Suppose we are given a partition of the set $\{1, 2, \ldots, m + n\}$ into two disjoint sets $A = \{a_1 < \cdots < a_m\}$ and $B = \{b_1 < \ldots < b_n\}$. We wish to determine the set A (and hence, B) by asking the minimum number of questions of the type: "Is $a_i > b_j$ or is $a_i < b_j$?" The answer to each question is known before the next question is asked. For each strategy S of asking questions, there is some choice of A and some set of answers which will require a maximum number of questions to determine A; we let $M(S)$ denote this maximum number of questions. Let $h(m, n)$ denote $\min_S M(S)$ where S ranges over all possible strategies which eventually determine A.

It is well-known (and easily shown) that $h(1, n) = 1 + [\log_2 n]$. In this paper we describe $h(2, n)$ as well as a related more general function. It will be seen, in particular, that the *least* value of n for which $h(2, n) = t \geqslant 2$ is exactly $[\frac{12}{7} \cdot 2^{(t-2)/2}]$ if t is even and $[\frac{17}{7} \cdot 2^{(t-3)/2}]$ if t is odd. The proofs of the assertions are rather lengthy and will not appear here.

2. A Generalization

For the case $m = 2$, write $A = \{\alpha < \beta\}$, $B = \{b_1 < \ldots < b_n\}$. Suppose that in addition to knowing the complete ordering within A and B we also have partial information concerning the ordering between A and B. Specifically, assume that for some i and j, $1 \leqslant i, j \leqslant n + 1$, we know $\alpha < b_i$ and $\beta > b_{n+1-j}$, where we use the convention that $\alpha < b_{n+1}$ indicates that we know nothing about the relative order of α and any element of B (with $\beta > b_0$ defined similarly). Let $f_n(i, j)$ denote $\min_S M(S)$ where S ranges over all strategies which eventually determine A. The main result of this paper is the determination of $f_n(i, j)$. Note that $f_n(n + 1, n + 1)$ is by definition equal to $h(2, n)$.

3. A Recursion for $f_n(i,j)$

We first note that $f_n(1,1) = 0$ for $n \geqslant 1$. Define $f_0(1,1) = 0$. Consider the situation which defines $f_n(i,j)$. That is, we know

$$\alpha < \beta, \quad b_1 < \ldots < b_n, \quad \alpha < b_i, \quad \beta > b_{n+1-j}.$$

Let S^* denote an optimal strategy for this situation, i.e., $M(S^*) = f_n(i,j)$. There are two possibilities for the first question of S^x.

(i) Suppose for some k the first question is "Is $\alpha \gtrless b_k$?". We can assume without loss of generality that $1 \leqslant k < i$ since no information can be gained if $k \geqslant i$. If the answer is "$\alpha < b_k$" then we are faced with the knowledge

$$\alpha < \beta, \quad b_1 < \ldots < b_n, \quad \alpha < b_k, \quad \beta > b_{n+1-j}.$$

By definition this situation requires $f_n(k,j)$ additional questions in order to determine A. On the other hand, if the answer is "$\alpha > b_k$", then we know

$$\alpha < \beta, \quad b_1 < \ldots < b_n, \quad \alpha > b_k, \quad \beta > b_{n+1-j}.$$

Note that now we can deduce $\beta > \alpha > b_k$ so that the elements $b_1, b_2, \ldots, b_k$ must just be the integers $1, 2, \ldots, k$. Hence, we are faced with the reduced situation

$$\alpha < \beta, \quad b_{k+1} < \ldots < b_n, \quad \alpha < b_i, \quad \beta > b_{n+1-j}.$$

Again, by definition, this requires $f_{n-k}(i-k,j)$ additional questions to determine A, where we make the convention that $f_m(x,y) = f_m(m+1,y)$ if $x \geqslant m+1$ with a similar convention holding if $y \geqslant m+1$.

Thus, we can write

$$f_n(i,j) \leqslant 1 + \max\left(f_n(k,j),\ f_{n-k}(i-k,j)\right).$$

(ii) Suppose for some k the first question of S^* is "Is $\beta \gtrless b_{n+1-k}$?". We can assume $1 \leqslant k < j$. Arguing as before we can conclude

$$f_n(i,j) \leqslant 1 + \max\left(f_n(i,k),\ f_{n-k}(i,j-k)\right).$$

Since in both (i) and (ii) the choice of k was arbitrary and since the first question of S^* must be either of type (i) or of type (ii) then we have the following recurrence:

$$f_n(i,j) = 1 + \min\left\{\min_{1 \leqslant k < i} \max\left(f_n(k,j),\ f_{n-k}(i-k,j)\right),\ \min_{1 \leqslant k < j} \max\left(f_n(i,k),\ f_{n-k}(i,j-k)\right)\right\} \tag{1}$$

for $n \geqslant 1$, $1 \leqslant i$, $j \leqslant n+1$, $(i,j) \neq (1,1)$, where $f_n(1,1) = 0$, for $n \geqslant 0$. It is this rather formidable looking recurrence which we shall solve.

4. The Structure of $f_n(i,j)$

We first make some preliminary remarks. It is easily checked that $f_n(i,j) = f_n(j,i)$. Further, if the value of α is determined by questions which never contain β, then since $\alpha < b_i$, we know that $h(1, i-1) = 1 + [\log_2(i-1)]$ questions are required (where we take $\log_2(0) \equiv -1$). Thus, by determining α and β independently we see that

$$f_n(i,j) \leqslant 2 + [\log_2(i-1)] + [\log_2(j-1)] \equiv R(i,j). \tag{2}$$

On the other hand, there are exactly $\left[ij - \binom{i+j-n-1}{2}\right]$ choices of A which satisfy the required conditions (where the binomial coefficient $\binom{x}{2}$ is taken to be 0 for $x < 2$). A standard argument of information theory shows

$$f_n(i,j) \geqslant \log_2\left[ij - \left(\frac{1+j-n-1}{2}\right)\right]. \tag{3}$$

A straightforward calculation yields

$$R(i,j) - \log_2\left[ij - \left(\frac{1+j-n-1}{2}\right)\right] < 3 \tag{4}$$

from which it follows, since $f_n(i,j)$ is an integer,

$$f_n(i,j) = R(i,j), \quad R(i,j) - 1 \quad \text{or} \quad R(i,j) - 2. \tag{5}$$

We shall call these values *regular*, *special* and *extraspecial* and denote them by R, S and X, respectively.†

As examples, we list the values of $f_n(i,j)$, $0 \leqslant n \leqslant 7$, in the form of square arrays in Table I. For a fixed value of n, the variables i, j each range over $1 \leqslant i, j \leqslant n+1 \equiv N$. The values of $f_n(1,1)$ and $f_n(1, n+1)$ are the upper left-hand and upper right-hand entries respectively. Thus $f_3(3,3) = 4$, $f_6(3,5) = 5$, etc.

Notice that all the entries in the tables for $N = 1$, 2, 4 and 8 are *regular*. This is not accidental.

Let us define certain subregions of the Nth array (the array which gives the values of $f_{N-1}(i,j)$). Write $N = 2^k + t$, $0 < t \leqslant 2^k$.

For $0 \leqslant r \leqslant k-1$, define the *rth region* of the array to be the set of coordinate values (i,j) for which $2^r < i \leqslant 2^{r+1}$, $2^k < j \leqslant N$. For $r = k$, the kth (or *critical*) *region* is defined by $2^k < i, j \leqslant N$.

† Thus, the knowledge that $\alpha < \beta$ can result in a savings of at most *two* questions.

TABLE I

$N = 1$

0

$N = 2$

0	1
1	2

$N = 3$

0	1	2
1	2	3
2	3	3

$N = 4$

0	1	2	2
1	2	3	3
2	3	4	4
2	3	4	4

$N = 5$

0	1	2	2	3
1	2	3	3	4
2	3	4	4	4
2	3	4	4	4
3	4	4	4	5

$N = 6$

0	1	2	2	3	3
1	2	3	3	4	4
2	3	4	4	5	5
2	3	4	4	5	5
3	4	5	5	5	5
3	4	5	5	5	5

$N = 7$

0	1	2	2	3	3	3
1	2	3	3	4	4	4
2	3	4	4	5	5	5
2	3	4	4	5	5	5
3	4	5	5	5	5	5
3	4	5	5	5	5	5
3	4	5	5	5	5	6

$N = 8$

0	1	2	2	3	3	3	3
1	2	3	3	4	4	4	4
2	3	4	4	5	5	5	5
2	3	4	4	5	5	5	5
3	4	5	5	6	6	6	6
3	4	5	5	6	6	6	6
3	4	5	5	6	6	6	6
3	4	5	5	6	6	6	6

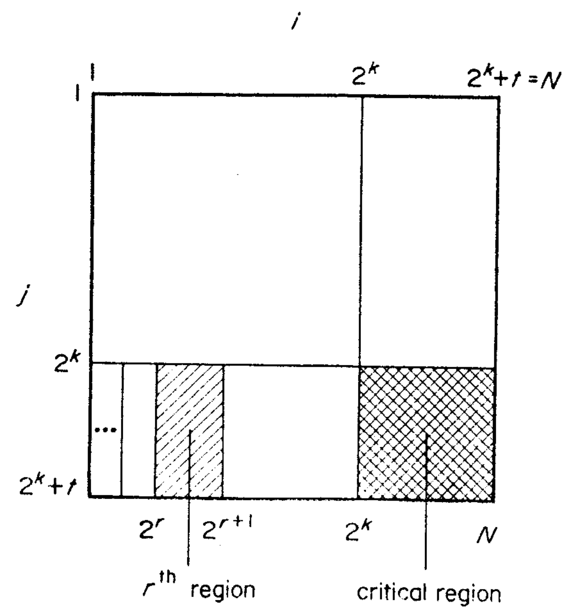

FIGURE 1

We next state a sequence of facts, whose proofs, as we have mentioned, will not be given here.

1. All entries $f_n(i,j)$, $1 \leqslant i, j \leqslant 2^k$, are *regular*, i.e., $f_n(i,j) = R(i,j)$.
2. All entries $f_n(1,j) = f_n(i,1)$, $1 \leqslant i,\ j \leqslant N$, are *regular*, i.e., $f_n(1,j) = 1 + [\log_2(j-1)]$.
3. The only *extraspecial* values occur in the critical region.
4. In general, the structure of the values of $f_n(i,j)$ in the rth region (and its symmetrical counterpart $f_n(j,i)$) is as follows, for $0 \leqslant r < k$:

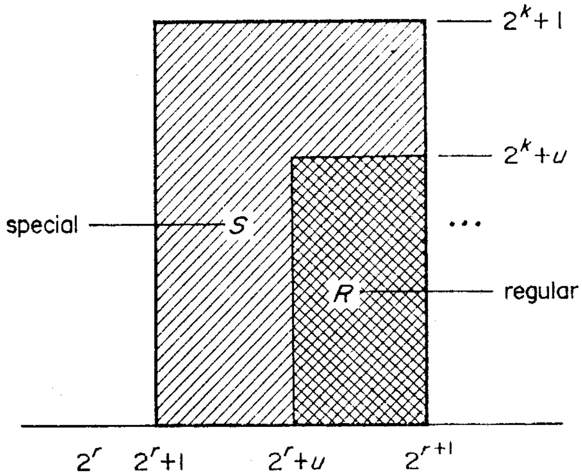

FIG. 2

There is a "strip" of *special* values of some uniform width u (possibly 0) which borders a rectangle of *regular* values as shown in Fig. 2. More precisely, for some $u \geqslant 0$, the values $f_n(i,j)$, $2^r + u < i \leqslant 2^{r+1}$, $2^k + u < j \leqslant N$, are regular. The remaining values in the rth region are special. For $r = k$, the same behaviour occurs in the critical region, both for a rectangle† of regular values bordered by a strip of special values, as well as a rectangle† of special values bordered by a strip of extraspecial values. The structure of the values in the critical region directly determines the structure of the values at these coordinates for larger values of N.

5. *Define* $F(N)$ by

$$F(N) = \begin{cases} X & \text{if all values in critical region are extraspecial,} \\ S & \text{if all values in critical region are special,} \\ u & \text{if there is a strip of } S(X) \text{ values of width } u \text{ which borders a} \\ & \text{rectangle of } R(S) \text{ values.} \end{cases}$$

† Which is in this region, in fact, a square.

We give a recursive definition for F from which the values of $f_n(i,j)$ can be immediately deduced. As usual, $[x]$ denotes the greatest integer $\leqslant x$. Let $F(1) = 0$, $F(2) = 0$, $F(3) = S$, $F(4) = 0$. For $N = n + 1 = 2^k + t$, $0 < t \leqslant 2^k$, $k \geqslant 2$,

$$F(N) = \begin{cases} X & \text{for } 0 < t \leqslant [\frac{3}{7} \cdot 2^{k-1}], \\ F(t) & \text{for } [\frac{3}{7} \cdot 2^{k-1}] < t \leqslant 2^{k-2}, \\ S & \text{for } 2^{k-2} < t \leqslant [\frac{5}{7} \cdot 2^k], \\ F(t - 2^{k-2}) + 2^{k-1} & \text{for } [\frac{5}{7} \cdot 2^k] < t \leqslant 3 \cdot 2^{k-2}, \\ 2^{k-1} & \text{for } 3 \cdot 2^{k-2} < t \leqslant [\frac{6}{7} \cdot 2^k], \\ F(t) & \text{for } [\frac{6}{7} \cdot 2^k] < t \leqslant 2^k. \end{cases} \tag{6}$$

Of course, the values of $F(N)$ for $[\frac{3}{7} \cdot 2^{k-1}] < t \leqslant 2^{k-2}$ correspond to the width of a strip of X values bordering a rectangle of S values. It is granted that this particular expression for $F(N)$ might not be the first thing that one would guess. It is true, however, that with sufficient patience one can indeed inductively establish (6) (and (7)).

Of course, in order to establish (6) one also must know $F^{(r)}(N)$, the width of the strip of S values which borders the rectangle of R values in the rth region, $1 \leqslant r < k$. These are recursively given by

$$F^{(r)}(N) = \begin{cases} S & \text{for } 0 < t \leqslant [\frac{5}{7} \cdot 2^r], \\ F(t - 2^{r-2}) + 2^{r-1} & \text{for } [\frac{5}{7} \cdot 2^r] < t \leqslant 3 \cdot 2^{r-2}, \\ 2^{r-1} & \text{for } 3 \cdot 2^{r-2} < t \leqslant [\frac{6}{7} \cdot 2^r], \\ F(t) & \text{for } [\frac{6}{7} \cdot 2^r] < t \leqslant 2^r, \\ 0 & \text{for } 2^r < t \leqslant 2^k. \end{cases} \tag{7}$$

6. It follows from (6) that $f_n(n + 1, n + 1)$ increases by 1 as N goes from $2^k + [\frac{3}{7} \cdot 2^{k-1}]$ to $2^k + [\frac{3}{7} \cdot 2^{k-1}] + 1$ and as N goes from $2^k + [\frac{5}{7} \cdot 2^k]$ to $2^k + [\frac{5}{7} \cdot 2^k] + 1$. Hence, the least value of n for which $h(2, n) = f_n(n + 1, n + 1) = t$ is exactly $[\frac{12}{7} \cdot 2^{(t-2)/2}]$ if t is even and $[\frac{17}{7} \cdot 2^{(t-3)/2}]$ if t is odd. Asymptotically, as n increases from 2^k to 2^{k+1}, the first 3/14 of the values of $h(2, n)$ are extraspecial, the next 1/2 are special and the final 2/7 are regular.

5. Concluding Remarks

The behaviour of $h(2, n)$ has recently been determined independently by Hwang and Lin (to appear). They did this without determining the general values of $f_n(i,j)$, by a combination of various bounds on $h(2, n)$ together with an ingenious explicit optimal strategy.

The determination of $h(3, n)$ would appear to be possible (but somewhat more difficult) using the techniques of this paper. The exact value of $h(m, n)$ for general m, however, would seem to require new ideas.

The author wishes to express his thanks to J. Evnine, F. K. Hwang, and S. Lin for stimulating discussions on this subject.

Reference

F. K. Hwang and S. Lin. An optimal algorithm for merging a linearly ordered set with an ordered set with two elements. *Acta Informatica.* To be published.