Randomized Selection in $n + C + o(n)$ comparisons

Alexandros V. Gerbessiotis

CS Department, New Jersey Institute of Technology, Newark, NJ 07102.

Constantinos J. Siniolakis

The American College of Greece, 6 Gravias St., Aghia Paraskevi, Athens, 15342, Greece.

Abstract

In this paper we present a randomized selection algorithm that with high probability $1 - 1/n^{\rho}$, for any constant $\rho > 1$ requires $n + C + o(n)$ comparisons to determine the C-th order statitistic of n keys thus matching a corresponding lower bound on the average number of comparisons required. Low order terms in the number of comparisons performed can also be reduced by extending the algorithm of Floyd and Rivest and analyzing its resulting performance more carefully.

Key words: **Keywords.** Randomized - Algorithms - Selection -

1 Introduction

We present a randomized selection algorithm that uses random oversampling and determines the C-th smallest of a set of n keys in $n+C+o(n)$ comparisons with probability at least $1 - 1/n^{\rho}$, for any constant $\rho > 0$. The selection problem that is, the computation of the C-th smallest key from a set of n keys has always been an interesting research topic [2,8,9]. Median finding in linear time is examined in [2,11] and lower bounds for various values of C are also presented in [2,8]. The minimum average number of comparisons required to find the C-th largest of n keys has also drawn considerable interest. A lower bound of $n + C - O(1)$ comparisons, $1 \leq C \leq \lceil n/2 \rceil$, is presented in [4], which matches the performance, within low order terms, of the algorithm in [5]. A sequential randomized algorithm for median finding that can be further extended to a parallel algorithm was proposed in [7]. The sequential version requires $3n/2 + O(n^{2/3} \log^{1/3} n)$ comparisons, with high probability. In this work we generalize the median finding result to the general selection problem. Our techniques provide tighter probabilistic bounds and optimal number of comparisons, within lower order terms.

We first present results on random sampling that will be used in the analysis of the selection algorithm that uses the technique of random oversampling. Random oversampling [10] results that generalize those first appearing in [7] for the special case of median finding are described using the notation of [6]. Let $X = \langle x_1, \ldots, x_N \rangle$ be a set of input keys ordered such that $x_i < x_{i+1}$, for $1 \leq i \leq N-1$. We are assuming that the keys are distinct because by appending to each key its index we can make them so. We are interested in finding the C-th smallest of the keys in set X. We assume, without loss of generality, that $C \leq \lceil N/2 \rceil$. Otherwise, we use symmetric arguments to find the $(N-C)$-th largest key. Let then be $c = C/N$. For integers $\lambda \geq 1$ and $s \geq 1$, let $Y = \langle y_1, \ldots, y_{2\lambda s-1} \rangle$ be a randomly chosen subset of $2\lambda s - 1 \leq N/2$ keys of X, also ordered so that $y_i < y_{i+1}$. The following three subsets are then formed for some positive integers μ_0, μ_2, and $t \geq 1$ such that $2c\lambda s > \mu_0 t/2$.

$$\begin{aligned}
X_0 &= \{x \mid x \in X \text{ and } x < y_{2c\lambda s - \mu_0 t/2}\}, \\
X_1 &= \{x \mid x \in X \text{ and } y_{2c\lambda s - \mu_0 t/2} < x < y_{2c\lambda s + \mu_2 t/2}\}, \\
X_2 &= \{x \mid x \in X \text{ and } y_{2c\lambda s + \mu_2 t/2} < x\}.
\end{aligned}$$

For the binomial coefficients, we assume that $\binom{i}{j} = 0$ if $j > i$. For $0 \leq i \leq 2$, let $n_i = |X_i|$ and let $p_i(j)$ be the probability that $n_i = j$. The following is established.

Claim 1 ([6]) *Probabilities $p_0(j)$ and $p_2(j)$ are given by the following expressions*

$$p_0(j) = \frac{\binom{j}{2c\lambda s - \mu_i t/2 - 1}\binom{N-j-1}{2(1-c)\lambda s + \mu_i t/2 - 1}}{\binom{N}{2\lambda s - 1}}, \ p_2(j) = \frac{\binom{j}{2(1-c)\lambda s - \mu_i t/2 - 1}\binom{N-j-1}{2c\lambda s + \mu_i t/2 - 1}}{\binom{N}{2\lambda s - 1}}.$$

We observe that the probability for $p_2(j)$ is that for $p_0(j)$ but with c replaced by $1-c$. The randomized selection algorithm we describe in the following section partitions the input key set into three sets as above. By establishing lower and upper bounds on the sizes of X_0 and X_2 we guarantee that the required statistic is located in set X_1. Our goal is thus to determine lower bounds on the size of set X_0, and symmetrically of X_2, by proving that the probability of having such a set of size less than $cN - f(N)$, for some function $f(N)$ of N, is negligible. We investigate bounds on $f(N)$ for this to hold. As the two cases for X_0 and X_2 are identically handled, we show the result by substituting $\mu_0 = \mu_2 = \mu$. Similarly, upper bounds on the size of set X_0, and symmetrically of X_2, can be established. Let $B = cN - f(N) - 1$. Particularly, we would like to show that the term $\sum_{j \le B} p_i(j)$, $i = 0, 2$, is sufficiently small for appropriate choices of $f(N)$. We deduce the following relationships used later for $n \ge N$. We note that the special case $c = 1/2$ was shown in [7].

Proposition 1 *Let $C \le \lceil N/2 \rceil$ and $c = C/N$. Let $\lambda, \mu \ge 1$, $s, t \ge 1$, $2\lambda s - 1 < N/2$, $n \ge N$, $2c\lambda s > \mu t/2$, $\mu t \ge 6$, and $\rho > 1$. If $f(N) = (N\ \mu t)/(2\ \lambda s)$ and $\mu t = 2\ \sqrt{3\rho\ 2c\ \lambda s\ \log n}$, then the probability that set X_0 is of size more than $cN - f(N) - 1$ is at least $1 - n^{1-\rho}$.*

Proof We first observe that from a combinatorial identity [9] we have

$$\sum_{j=2c\lambda s-\mu t/2-1}^{N-2(1-c)\lambda s-\mu t/2} \binom{j}{2c\lambda s - \mu t/2 - 1}\binom{N-j-1}{2(1-c)\lambda s + \mu t/2 - 1} = \binom{N}{2\lambda s - 1}.$$

Let $B = cN - f(N) - 1$. We use a combinatorial interpretation argument, as in [6], to argue that the sum $\sum_{j \le B} p(j)$ for $j \le B$ is the number of ways we can choose $2\lambda s - 1$ objects so that among them, the object with the $(2c\lambda s - \mu t/2)$-th highest label is labeled at most B. In order to establish an upper bound for $\sum_{j \le B} p(j)$ we approximate $p(j)$ by a binomial term and use bounds on the right tail of the binomial distribution. Let $q = (2\lambda s - 1)/N$. The probability that in N Bernoulli trials, each with a probability q of success we get $2c\lambda s - \mu t/2 - 1$ successes among the first j, a success at the $(j+1)$-st trial, and $2(1-c)\lambda s + \mu t/2 - 1$ successes in the remainder is

$$\begin{aligned} r(j) = &\binom{j}{2c\lambda s - \mu t/2 - 1} q^{2c\lambda s - \frac{\mu t}{2} - 1}(1-q)^{j - 2c\lambda s + \frac{\mu t}{2} + 1} \\ &q\binom{N-j-1}{2(1-c)\lambda s + \mu t/2 - 1} q^{2(1-c)\lambda s + \frac{\mu t}{2} - 1}(1-q)^{N-j-2(1-c)\lambda s - \frac{\mu t}{2}}. \end{aligned}$$

Let A be the ratio of $r(j)$ over $p(j)$. The following then holds.

$$A = \frac{r(j)}{p(j)} = \binom{N}{2\lambda s - 1} q^{2\lambda s - 1}(1-q)^{N - 2\lambda s + 1}.$$

The following inequality is shown in [3]: $exp(-1/(6c))(1/\sqrt{2\pi abd})\ \left(a^a b^b\right)^{-d} \le \binom{d}{c}$. By substituting for $d = N$, $c = 2\lambda s - 1$, $a = q = (2\lambda s - 1)/N$, and $b = 1 - a$ we get the following result that provides a lower bound for A.

$$\frac{e^{-\frac{1}{6(2\lambda s - 1)}}\ q^{-2\lambda s + 1}\ (1-q)^{-N + 2\lambda s - 1}}{\sqrt{2\pi(2\lambda s - 1)(1 - (2\lambda s - 1)/N)}} \le \binom{N}{2\lambda s - 1},$$

$$A \geq \frac{e^{-\frac{1}{6(2\lambda s-1)}}}{\sqrt{2\pi(2\lambda s-1)(1-(2\lambda s-1)/N)}}. \tag{1}$$

From the lower bound on A, we obtain the following result.

$$\sum_{j\leq B} p(j) \leq \frac{q}{A}\sum_{j\leq B}\binom{j}{2c\lambda s-\mu t/2-1} q^{2c\lambda s-\mu t/2-1}(1-q)^{j-2c\lambda s+\mu t/2+1} \tag{2}$$

The goal is to establish an upper bound for the term below.

$$\sum_{j\leq B}\binom{j}{2c\lambda s-\mu t/2-1} q^{2c\lambda s-\mu t/2-1}(1-q)^{j-2c\lambda s+\mu t/2+1}.$$

We employ generating functions. The term

$$\binom{j}{2c\lambda s-\mu t/2-1} q^{2c\lambda s-\mu t/2-1}(1-q)^{j-2c\lambda s+\mu t/2+1}$$

is the coefficient of $x^{2c\lambda s-\mu t/2-1}$ in the polynomial $((1-q)+qx)^j$. Therefore, the sum over all those terms with $j \leq B$ is the coefficient of $x^{2c\lambda s-\mu t/2-1}$ in the following sum.

$$\sum_{j\leq B}((1-q)+qx)^j = \frac{((1-q)+qx)^{B+1}}{-q(1-x)} + \frac{1}{q(1-x)}.$$

The coefficient of $x^{2c\lambda s-\mu t/2-1}$ in this expression is the following one.

$$\frac{1}{q}\left(1-\sum_{i=0}^{2c\lambda s-\frac{\mu t}{2}-1}\binom{B+1}{i}q^i(1-q)^{B+1-i}\right) = \frac{1}{q}\sum_{i=2c\lambda s-\frac{\mu t}{2}}^{B+1}\binom{B+1}{i}q^i(1-q)^{B+1-i}.$$

The sum in the term above gives the probability of having at least $2c\lambda s-\mu t/2$ successes in $B+1$ independent Bernoulli trials with probability of success q. Chernoff's upper bound [1] on the probability of having at least $(1+\beta)(B+1)q$ successes in such an experiment gives an upper bound of $e^{-(\beta^2(B+1)q)/3}$. We set $2c\lambda s-\mu t/2 = (1+\beta)(B+1)q$ in this bound and get $e^{-(2c\lambda s-\mu t/2-(B+1)q)^2/(3(B+1)q)}$. We choose $f(N)$ so that $f(N) = (N\ \mu t)/(2\ \lambda s)$. A lower bound for $2c\lambda s - \mu t/2-(B+1)q$ and an upper bound for $(B+1)q$ are derived similarly.

$$2c\lambda s - \mu t/2 - (B+1)q \geq \mu t/2 - (\mu t)/(2\lambda s) + c \geq \mu t/2 - (\mu t)/(2\lambda s).$$

$$(B+1)q \leq (2c\lambda s - \mu t) + (\mu t)/(2\ \lambda s) - c \leq (2c\lambda s - \mu t) + (\mu t)/(2\ \lambda s).$$

We thus get for the sum of interest.

$$\sum_{j\leq B} \binom{j}{2c\lambda s - \frac{\mu t}{2} - 1} q^{2c\lambda s - \frac{\mu t}{2} - 1}(1-q)^{j-2c\lambda s + \frac{\mu t}{2} + 1} \leq e^{-\frac{(2c\lambda s - \mu t/2 - (B+1)q)^2}{(3(B+1)q)}}. \quad (3)$$

By way of equations (1), (2) and (3) we have that

$$\sum_{j\leq B} p(j) \leq \sqrt{2\pi(2\lambda s - 1)}\ e^{\frac{1}{6(2\lambda s - 1)}}\ e^{-\frac{(2c\lambda s - \mu t/2 - (B+1)q)^2}{3(B+1)q}}.$$

From the lower bound of $2c\lambda s - \mu t/2 - (B+1)q$ derived earlier and the corresponding upper bound for $(B+1)q$ we get that, since $c \leq 1$, and for $\mu t \geq 6 \geq 4c + 2c \geq 4c + \mu t/(2\lambda s)$,

$$e^{-(2c\lambda s - \mu t/2 - (B+1)q)^2/(3(B+1)q)} \leq e^{-(1/3)(\mu t/2)^2/(2c\lambda s)}.$$

Substitution for $\mu t = 2\ \sqrt{3\rho\ 2c\ \lambda s\ \log n}$ and $\sqrt{2\pi(2\lambda s - 1)}\ e^{\frac{1}{6(2\lambda s - 1)}} < n$, gives the desired result. ■

We can establish a similar bound for X_2 noting the symmetry between $p_0(j)$ and $p_2(j)$. Variable μ is μ_2 in this case.

Proposition 2 *Let $C \leq \lceil N/2 \rceil$ and $c = C/N$. Let $\lambda, \mu \geq 1$, $s, t \geq 1$, $2\lambda s - 1 < N/2$, $n \geq N$, $2(1-c)\lambda s > \mu t/2$, $\mu t \geq 6$, and $\rho > 1$. If $f(N) = (N\ \mu t)/(2\ \lambda s)$ and $\mu t = 2\ \sqrt{3\rho\ 2(1-c)\ \lambda s\ \log n}$, then the probability that set X_2 is of size more than $(1-c)N - f(N) - 1$ is at least $1 - n^{1-\rho}$.*

Proof The proof is identical to the proof of Proposition 1. Since in $p_2(j)$, $1-c$ appears in place of c in $p_0(j)$ the only changes in the statement are due to this fact. For the error probability bound derivation, instead of requiring the previous condition $\mu t \geq 6 \geq 4c + 2c \geq 4c + \mu t/(2\lambda s)$, we require that the following one is true $\mu t \geq 6 \geq 4(1-c) + 2c \geq 4(1-c) + \mu t/(2\lambda s)$, which, however, leads to the same lower bound for μt. ■

We can similarly establish upper bounds for the sizes of the sets X_0 and X_2.

Proposition 3 *Let $C \leq \lceil N/2 \rceil$ and $c = C/N$. Let $\lambda, \mu \geq 1$, $s, t \geq 1$, $2\lambda s - 1 < N/2$, $n \geq N$, $2c\lambda s > \mu t/2$, and $\rho > 1$. If $\mu t \geq 2\sqrt{2\rho\, 2c\, \lambda s\, \log n}$ then the probability that set X_0 is of size at most $cN - 1$ is at least $1 - n^{1-\rho}$.*

Proof The proof is a generalization of the result shown in [7] and its derivation is similar to the one in the proof of Proposition 1 provided that we set $f(N) = 0$ there, i.e. set B to $B = cN$. The probability bound we obtain is

$$\sum_{j \geq B} p_0(j) \leq \sqrt{2\pi(2\lambda s - 1)}\; e^{\frac{1}{6(2\lambda s - 1)}}\; e^{-\frac{(Bq - 2c\lambda s + \mu t/2 + 1)^2}{2Bq}}.$$

The nominator $(Bq - 2c\lambda s + \mu t/2 + 1)$ has $\mu t/2$ as a lower bound, the denominator $2Bq$ has $2\lambda sc$ as an upper bound. The choice of μt leads to the desired result. ■

For X_2 a symmetric result can be shown, where μ is to be interpreted as μ_2.

Proposition 4 *Let $C \leq \lceil N/2 \rceil$ and $c = C/N$. Let $\lambda, \mu \geq 1$, $s, t \geq 1$, $2\lambda s - 1 < N/2$, $n \geq N$, $2(1-c)\lambda s > \mu t/2$, and $\rho > 1$. If $\mu t \geq 2\sqrt{2\rho\, 2(1-c)\, \lambda s\, \log n}$ then the probability that set X_2 is of size at most $cN - 1$ is at least $1 - n^{1-\rho}$.*

The following bounds on the sizes of X_0 and X_2 are derived for $s = t = 1$.

Claim 2 *Let $C \leq \lceil N/2 \rceil$ and $c = C/N$. Let $\lambda \geq 1$, $2\lambda - 1 < N/2$, $n \geq N$, and $\rho > 1$. Let $\mu_0 \geq 2\sqrt{2\rho\, 2c\, \lambda\, \log n} \geq 1$ and $2c\,\lambda > \mu_0/2$. Then, with probability at least $1 - n^{1-\rho}$, set X_0 is of size at most $C - 1$.*

Proof The proof follows directly from proposition 3 with $s = t = 1$. We only note that the failure probability bound $n^{1-\rho}$ is an overestimate of a tighter one $n^{\eta-\rho}$, where η is any constant greater than 0.5 and smaller than 1.0. ■

Claim 3 *Let $C \leq \lceil N/2 \rceil$ and $c = C/N$. Let $\lambda \geq 1$, $2\lambda - 1 < N/2$, $n \geq N$, and $\rho > 1$. Let $\mu_2 \geq 2\sqrt{2\rho\, 2(1-c)\, \lambda\, \log n} \geq 1$ and $2(1-c)\,\lambda > \mu_2/2$. Then, with probability at least $1 - n^{1-\rho}$, set X_2 is of size at most $N - C - 1$.*

Proof The proof follows directly from proposition 4 with $s = t = 1$. ■

Claim 4 *Let $C \leq \lceil N/2 \rceil$ and $c = C/N$. Let $\lambda \geq 1$, $2\lambda - 1 < N/2$, $n \geq N$, and $\rho > 1$. Let $2c\lambda > \mu_0/2$ and $\mu_0 = 2\sqrt{3\rho\, 2c\, \lambda\, \log n} \geq 3$. Then, with probability at least $1 - n^{1-\rho}$, set X_0 is of size at least $cN - N\sqrt{3\rho\, 2c\, \log n}/\sqrt{\lambda}$.*

Proof The claim is a direct consequence of proposition 1 by substituting for $s = t = 1$ and $\mu = 2\sqrt{3\rho\, 2c\, \lambda\, \log n}$ for $f(N)$. ■

Claim 5 *Let* $C \leq \lceil N/2 \rceil$ *and* $c = C/N$. *Let* $\lambda \geq 1$, $2\lambda - 1 < N/2$, $n \geq N$, *and* $\rho > 1$. *Let* $2(1-c)\lambda > \mu_2/2$ *and* $\mu_2 = 2\ \sqrt{3\rho\, 2(1-c)\ \lambda\ \log n} \geq 3$. *Then, with probability at least* $1 - n^{1-\rho}$, *set* X_2 *is of size at least* $N - cN - N\sqrt{3\rho\, 2(1-c)\ \log n}/\sqrt{\lambda}$.

Proof The claim is a direct consequence of proposition 2 by substituting for $s = t = 1$ and $\mu = 2\ \sqrt{3\rho\, 2(1-c)\ \lambda\ \log n}$ for $f(N)$. ■

2 Randomized Selection

The randomized selection algorithm derives from ideas in [5] and extends the results presented in [7]. The general formulation of the algorithm is as follows. A random sample of the input is initially chosen, and two keys within this sample are selected so that the C-th smallest key is between them.

begin SELECT (X, C)
- **let** $c = C/n$, $\lambda = n^{2/3}(3\rho \log n)^{1/3}((\sqrt{2c} + \sqrt{2(1-c)})/2)^{2/3}$,
  such that $2\lambda - 1 < n/2$;
- **let** $\mu_0 = 2\sqrt{3\rho\ 2c\ \lambda \log n}$, $\mu_2 = 2\sqrt{3\rho\ 2(1-c)\ \lambda \log n}$;
- **select** randomly a sample $S = \{y_1, y_2, \ldots, y_{2\lambda-1}\}$ of size $2\lambda - 1$;
- denote by $\langle y_1^*, y_2^*, \ldots, y_{2\lambda-1}^* \rangle$ the sorted sequence S ;
- **if** $(c \geq 2\rho \log n/\lambda)$ **then**
  - **let** $m = y^*_{2c\lambda - \mu_0/2}$, $M = y^*_{2c\lambda+\mu_2/2}$;
- **else**
  - **let** $m = -\infty$, $M = y^*_{2c\lambda+\mu_2/2}$;
- **let** $r = l = 0$, $S = R_r = R_l = \{\}$;
- **for** $i = 1$ **to** n
- **do if** $(x_i > M)$ **then**
  - **let** $r = r + 1$, $R_r = R_r \bigcup \{x_i\}$;
  - **elseif** $(x_i < m)$ **then**
    - **let** $l = l + 1$, $R_l = R_l \bigcup \{x_i\}$;
  - **else**
    - **let** $S = S \bigcup \{x_i\}$;
- **if** $(l < C$ and $r < n - C)$ **then**
  - **return** SELECT_DET$(S, C - l)$;
- **else**
  - **reexecute** procedure SELECT ;

Fig. 1. Procedure SELECT

Theorem 1 *For any constant $\rho > 1$, there exists an algorithm that, with probability at least $1 - O(n^{1-\rho})$, determines the C-th smallest, $C \leq \lceil n/2 \rceil$, of n keys within $n + C + o(n)$ comparisons.*

Proof We note that if $C > \lceil n/2 \rceil$, by reversing the meaning of the inequalities in procedure Select, we can employ it to determine the $(n - C)$-th largest key of the input key set, which is the C-th smallest key of the set. Procedure Select that determines the C-th smallest key of a set X of size n is given in pseudocode in figure 1. A high-level description of the algorithm is also given below.

Stage 1: [Lines 1-8 of Select in figure 1] Let $c = C/n$ and set λ such that $\lambda = n^{2/3}(3\rho \log n)^{1/3}((\sqrt{2c} + \sqrt{2(1-c)})/2)^{2/3}$, so that $2\lambda - 1 < n/2$. Variables μ_0 and μ_2 are set to $\mu_0 = 2\sqrt{3\rho\, 2c\ \lambda\ \log n}$ and $\mu_2 = 2\sqrt{3\rho\, 2(1-c)\ \lambda\ \log n}$ respectively. A random sample of size $2\lambda - 1$ is selected. If $c \geq 2\rho \log n/\lambda$, the $(2c\lambda - \mu_0/2)$-th and $(2c\lambda + \mu_2/2)$-th keys of the sample are determined using the linear time selection algorithm in [2]. Otherwise, the $(2c\lambda + \mu_2/2)$-th key of the sample is only determined using the algorithm in [2]. Let $m = y_{2c\lambda - \mu_0/2}$ (resp. $m = -\infty$) and $M = y_{2c\lambda + \mu_2/2}$ be the so identified keys for the two cases respectively.

Stage 2. [Lines 9-16 of Select in figure 1] Let $l = r = 0$ and $S = \{\}$. For each key x in X, x is compared to M first. If x is larger than M, then $r = r + 1$. If x is smaller than M, then x is compared to m. If it is found smaller than m, then $l = l + 1$. Otherwise, $S = S \cup \{x\}$.

Stage 3. [Lines 17-20 of Select in figure 1] If $l < C$ and $r < n - C$, the $(C - l)$-th smallest key of set S is determined using the linear-time selection algorithm in [2]. Otherwise failure is declared and the algorithm is reexecuted.

We distinguish two cases in analyzing the performance of Select.

Case 1: $(c \geq 3\rho \log n/\lambda)$. Stage 1 of the algorithm, by the choice of λ requires $O(\lambda)$ comparisons. By claims 2, 3, 4 and 5, for $N = n$, at least $C - n\sqrt{3\rho\, 2c\ \log n}/\sqrt{\lambda}$ and at most $C - 1$ keys have value less than m, and at least $n - C - n\sqrt{3\rho\, 2(1-c)\ \log n}/\sqrt{\lambda}$ and at most $n - C - 1$ keys have values larger than M, with probability at least $1 - n^{1-\rho}$, for any constant $\rho > 1$. With the stated probability the number of comparisons performed at stage 2 is at most $(n - C) + 2(C + n\sqrt{3\rho\, 2(1-c)\ \log n}/\sqrt{\lambda})$ which is $n + C + 2n\sqrt{3\rho\, 2(1-c)\ \log n}/\sqrt{\lambda}$. The size of set S is also bounded above by $(n\sqrt{3\rho\ \log n}/\sqrt{\lambda})(\sqrt{2c} + \sqrt{2(1-c)})$. This latter term is 2λ. Stage 3 then, by way of the algorithm in [2] and the value of λ requires $O(n^{2/3}(6\rho\ \lg n)^{1/3}(\sqrt{2c} + \sqrt{2(1-c)})^{2/3})$ comparisons. An upper bound of $n + C + O(n^{2/3}(6\rho \lg n)^{1/3}(\sqrt{2c} +$

$\sqrt{2(1-c)})^{2/3})$ comparisons is proved with probability at least $1 - n^{1-\rho}$, for any constant $\rho > 1$.

Case 2: $(c < 3\rho \log n/\lambda)$. Stages 1 and 2 require the same number of comparisons as for the previous case. In stage 3 the number of keys with value larger than M is at least $n - C - n\sqrt{3\rho\, 2(1-c)\ \log n}/\sqrt{\lambda}$ and therefore, $|S|$ is at most $C + n\sqrt{3\rho\, 2(1-c)\ \log n}/\sqrt{\lambda}$ with probability at least $1 - n^{1-\rho}$, for any constant $\rho > 1$. Since $c < 3\rho \log n/\lambda$, with the stated probability, the size of S is bounded above by $3\rho n \log n/\lambda + n\sqrt{3\rho\, 2(1-c)\ \log n}/\sqrt{\lambda}$. The first term is thus insignificant and the same upper bound as for the previous case can be claimed for the total comparison count. ■

References

[1] D. Angluin and L. G. Valiant. Fast probabilistic algorithms for hamiltonian circuits and matchings. *Journal of Computer and System Sciences*, 18:155-193, 1979.

[2] M. Blum, R. W. Floyd, V. R. Pratt, R. L. Rivest and R. E. Tarjan. Time bounds for selection. *Journal of Computer and System Sciences*, 7:448-461, 1973.

[3] B. Bollobás. *Random Graphs.* Academic Press, New York, 1984.

[4] W. Cunto and J. I. Munro. Average case selection. In *Proceedings of the 16-th Annual ACM Symposium on Theory of Computing*, pp. 369-375, 1984.

[5] R. W. Floyd and R. L. Rivest. Expected time bounds for selection. *Communications of the ACM*, 18:165-172, 1975.

[6] A. V. Gerbessiotis and L. G. Valiant. Direct Bulk-Synchronous Parallel Algorithms. *Journal of Parallel and Distributed Computing*, 22:251-267, 1994.

[7] A. V. Gerbessiotis and C. J. Siniolakis. Deterministic sorting and randomized median finding on the BSP Model. In *Proceedings of the 8-th ACM Symposium on Parallel Algorithms and Architectures*, June 1996.

[8] D. G. Kirkpatrick. A unified lower bound for selection and set partitioning problems. *Journal of the ACM*, 28:150-165, 1981.

[9] D. E. Knuth. *The Art of Computer Programming. Volume I: Fundamental Algorithms.* Addison-Wesley, Reading, 1969.

[10] H. J. Reif and L. G. Valiant. A logarithmic time sort for linear size networks. *Journal of the ACM,* 34:60-76, January 1987.

[11] A. Schonhage, M. Paterson, and N. Pippenger. Finding the median. *Journal of Computer and System Sciences*, 13:184-199, 1976.