1. INTRODUCTION

1.1 The Problem

The classical problem of selecting the k-th smallest element of a set F drawn from a totally ordered set has been extensively studied in serial and parallel environments. In a distributed context, it has different formulations and complexity measures.

A communication network of size \( d \) is a set \( S = \{S_0, \ldots, S_{d-1}\} \) of sites, where each site has a local non-shared memory and processing capabilities. In the point-to-point model, associated with \( S \) is a set \( L \subseteq S \times S \) of direct communication lines between sites; if \((S_i, S_j) \in L\), \( S_i \) and \( S_j \) are said to be neighbours. Sites communicate by sending messages; a message can only be sent to and received from a neighbour. The couple \( G = (S, L) \) can be thought of as an undirected graph; hence, graph-theoretical notation can be employed in the design and analysis of distributed algorithms in the point-to-point model.

A file of cardinality \( N \) is a set \( F = \{f_1, \ldots, f_N\} \) of records, where each record \( f \in F \) contains a unique key \( k(f) \) drawn from a totally ordered set \( F \); in the following, \( f_i > f_j \) will denote that \( k(f_i) > k(f_j) \).

A distribution of \( F \) on \( S \) is a d-tuple \( X = \langle X_0, \ldots, X_{d-1} \rangle \) where \( X_i \subseteq F \) is a subfile stored at site \( S_i \), \( |X_i| \leq c \), \( X_i \cap X_j = \emptyset \) for \( i \neq j \), and \( \bigcup_i X_i = F \).

Order-statistics queries about \( F \) can be originated at any site and will activate a query resolution process at that site. Since only a subset of \( F \) is available at each site, the resolution of a query will in general require the cooperation of several (possibly all) sites according to some predetermined algorithm. Since local processing time is usually negligible when compared with transmission and queueing delays, the goal is to design resolution algorithms which minimize the amount of communication activity rather than the amount of processing activity.

The distributed selection problem is the general problem of resolving a query for locating the K-th smallest element of \( F \). The tuple \( \langle N, K, N[0], \ldots, N[d-1] \rangle \) is called the problem configuration, and \( \Delta = \min\{K, N-K+1\} \) is called the problem size, where \( N[i] \) is the cardinality of \( X_i \). Any efficient solution to this problem can be employed as a building block for a distributed sorting algorithm [11].
The complexity of this problem (i.e., the number of communication activities required to resolve an order-statistics query) depends on many parameters, including the number $d$ of sites, the size $N$ of the file, the number $N[i]$ of elements stored at site $S_i$, the rank $K$ of the element being sought, the topology of the network. This work deals with applications for which the size of the file is much greater than the number of sites; i.e., $N \gg d$.

1.2 An Historical Prospective

The current definition of the distributed selection problem is the result of a synthesis of two distinct types of investigations carried out in the past, each having its own motivations and assumptions.

The first type of investigations on the distributed selection problem and the related ranking problem [5,6,19] were generalizations of the studies on the minimum-finding (or election) problem in the point-to-point model, where it was assumed that each site contained one element. Several solutions have been presented [3,8,17]; in all these investigations, the focus was on the worst-case complexity; the only exception is the distributed translation by Shrira, Francez and Rodeh [17] of the well-known serial algorithm for random selection.

A different motivation came from the investigations by Yao [18] and Abelson [1] on the amount of communication needed to evaluate a smooth Boolean function whose arguments were stored at two different sites. The case where the arguments of the function are elements from a totally ordered set, and the result of the function is the median of the arguments was independently studied by Rodeh [9] and Santoro & Sidney [14] who proved a $\Theta(\log N)$ bound on the problem for $d=2$; the constant in the bound was later improved [2]. The generalization to the case $d>2$ was then the object of several investigations in the so-called shout-echo model (a topology-independent model allowing broadcast-type primitives) [7,10,12,14-16]. Almost all solutions obtained in this model are easily convertible to solutions in the more common point-to-point model (used in the other class of investigations). Also these investigations have focused on the worst-case complexity.

One important fact learned in the investigations with the shout-echo model is about the nature of the distributed selection problem itself. Since it is assumed that the cost of local processing is negligible, is it reasonable to assume that the elements stored at a site are sorted; thus, the distributed selection problem can be seen as the distributed equivalent of the serial problem of selection in an array with sorted columns where each column correspond to a subfile. Based on this observation, a distributed version of the optimal serial algorithm by Frederickson and Johnson [4] was developed leading to a $\Theta(d \log(k/d))$ algorithm for complete and star graphs in the point-to-point model and to a $\Theta(\log(k/d))$ algorithm in the shout-echo model [16]. Again, these bounds apply to the worst-case complexity.

In this paper, the expected communication complexity of the distributed selection problem is analysed.