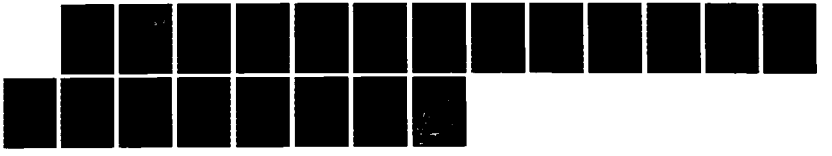
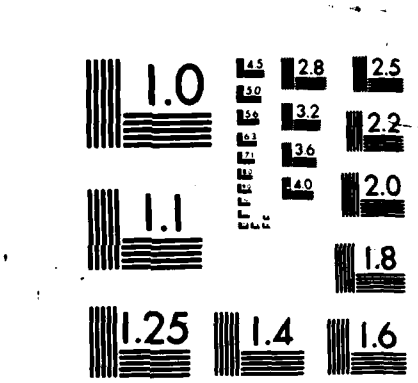
AD-A172 224 DISTRIBUTED CONTROL IN COMPUTER NETWORKS AND CROSS-SECTIONS OF COLORED MU.. (U) MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR COMPUTER SCIENCE.. 1/1

UNCLASSIFIED E KRANAKIS ET AL. APR 86 MIT/LCS/TM-304 F/G 9/2 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1
8.0
9.0
10.0

LABORATORY FOR
COMPUTER SCIENCE

MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

MIT/LCS/TM-304

DISTRIBUTED CONTROL IN COMPUTER NETWORKS AND CROSS-SECTIONS OF COLORED MULTIDIMENSIONAL BODIES

EVANGELOS KRANAKIS
PAUL M. B. VITANYI

APRIL 1986

545 TECHNOLOGY SQUARE, CAMBRIDGE, MASSACHUSETTS 02139

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER MIT/LCS/TM-304	2. GOVT ACCESSION NO. AD-A172 224	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Distributed control in computer networks and cross-sections of colored multidimensional bodies.		5. TYPE OF REPORT & PERIOD COVERED Interim research April 1986
		6. PERFORMING ORG. REPORT NUMBER MIT/LCS/TM-304
7. AUTHOR(s) Evangelos Kranakis and Paul M. B. Vitanyi		8. CONTRACT OR GRANT NUMBER(s) N00014-83-K-0125
9. PERFORMING ORGANIZATION NAME AND ADDRESS MIT Laboratory for Computer Science 545 Technology Square Cambridge, MA 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS DARPA/DOD 1400 Wilson Boulevard Arlington, VA 22209		12. REPORT DATE April 1986 13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) ONR/Department of the Navy Information Systems Program Arlington, VA 22217		15. SECURITY CLASS. (of this report) unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Unlimited

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Distributed match-making, computer network, distributed control, name server, mutual exclusion, colored body, measure

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The number of messages to match a pair of processes in a multiprocessor network with mobile processes is a measure for the cost of setting up temporary communication between processes. We establish lower bounds on the average number of point-to-point transmissions between any pair of nodes in this context. The present analysis allows for the possibility of multiple transmissions (as opposed to a single one)between any two nodes, and also for

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

the possibility of multiple queries(as opposed to the two, i.e. post and a single query considered before). Applications of the results include lower bounds on the number of messages for distributed s-matching, that is, matching a group of s processes, and distributed s-mutual exclusion, that is, s-1 processes may enter a critical section simultaneously, but s processes may not, for $s \geq 2$. The idea of the proof of the combinatorial result needed for this analysis is further extended to obtain a lower bound on the average number of colors occurring in random cross-sections of colored, multidimensional bodies in terms of the total (multi-dimensional) volume of each color in the whole body.

Distributed Control in Computer Networks and Cross-Sections of Colored Multidimensional Bodies

Evangelos Kranakis

Centrum voor Wiskunde en Informatica
Amsterdam, The Netherlands

Paul M. B. Vitányi

Massachusetts Institute of Technology
Laboratory for Computer Science
Cambridge, Massachusetts 02139
and
Centrum voor Wiskunde en Informatica
Amsterdam, The Netherlands

April, 1986

ABSTRACT

The number of messages to match a pair of processes in a multiprocessor network with mobile processes is a measure for the cost of setting up temporary communication between processes. We establish lower bounds on the average number of point-to-point transmissions between any pair of nodes in this context. The present analysis allows for the possibility of multiple transmissions (as opposed to a single one) between any two nodes, and also for the possibility of multiple queries (as opposed to the two, i.e. post and a single query considered before). Applications of the results include lower bounds on the number of messages for distributed s-matching, that is, matching a group of s processes, and distributed s-mutual exclusion, that is, $s-1$ processes may enter a critical section simultaneously, but s processes may not, for $s \geq 2$. The idea of the proof of the combinatorial result needed for this analysis is further extended to obtain a lower bound on the average number of colors occurring in random cross-sections of colored, multidimensional bodies in terms of the total (multidimensional) volume of each color in the whole body.

Key Words and Phrases: Distributed Match-Making, Computer Network, Distributed Control, Name Server, Mutual Exclusion, Colored Body, Measure.

The work of the first author was supported in part by the Computer Science Department of the University of Amsterdam. The work of the second author was supported in part by the Office of Naval Research under Contract N00014-85-K-0168, by the Office of Army Research under Contract DAAG29-84-K-0058, by the National Science Foundation under Grant DCR-83-02391, and by the Defense Advanced Research Projects Agency (DARPA) under Grant N00014-83-K-0125.

1. Introduction

The design objectives of the *Amoeba* distributed operating system (described in [9]) motivated (see [5], [6]) the design and analysis of a mathematical model for the so called *name-server* mechanism in a distributed system with mobile processes (and objects, henceforth subsumed under processes). The name server, that is, a system that translates names of services into locations in the network, forms a central part of the design of many distributed operating systems for computer networks and multiprocessor systems. The implementation has been approached in various ways, from centralized directories which are vulnerable to host processor crashes [8], to methods which maintain a tree of forwarding addresses in the network for each mobile process [7]. This latter method is not more robust, and in general much less efficient, than the class of methods we consider below.

More generally, we address the problem of matching mobile processes in a multiprocessor environment without centralized control. We call this *distributed match-making* (see [6]). Various issues in distributed control can be thought of in terms of the distributed match-making paradigm. One of them is the name-server, another one is mutual exclusion.

1.1. The Name-Server

A set of named processes with no permanent addresses residing at the nodes of a network wish to set up communication, when needed, among themselves. Let N be the set of nodes (i.e., processors) of such a network. The network is a communication graph with two-way noninterfering communication channels between directly connected nodes. It is assumed that the nodes communicate only by messages and do not share memory. An error-free underlying communications network supports the message transfers in which the delivery time may vary but messages between two nodes are delivered in the order sent. Each of these processes is considered both a potential server (i.e. it can offer a service) as well as a potential customer (i.e. it may request a service). Let a process p reside at a host node $h(p)$. Since processes may migrate, die or be created, $h(p)$ can change, become empty or nonempty. Here we make the simplifying assumption that for the segment of time of the actual match-making instance, the process/processor allocation does not change. Location of services by the processes is achieved by the following procedure. Each server s selects a set $P(s)$ of nodes and posts at these nodes the availability of the service it offers and the address $h(s)$ where it resides. (Each node in $P(s)$ stores this information in its individual *cache*.) When a client c wants to request a service it selects a set $Q(c)$ of nodes and queries each node in $Q(c)$ for the required service. When $P(s) \cap Q(c)$ is not empty the node (or any node) in $P(s) \cap Q(c)$ will be able to return a message to c stating the address $h(s)$ at which the service is available (recall that this information is already stored in the caches of all the nodes in $P(s)$). For example, a *centralized* name-server corresponds to

$$P(s) = \{x\},\ Q(c) = \{x\},$$

broadcasting corresponds to

$$P(s) = \{h(s)\},\ Q(c) = N,$$

while what we may call *sweeping* corresponds to

$$Q(c) = \{h(c)\},\ P(s) = N,$$

for all servers s and clients c with $h(s), h(c) \in N$ and some $x \in N$. Another example is the *Manhattan network*. The set N of nodes consists of pairs (i,j), with $i=1,...,m$, $j=1,...,n$. For all $(i,j) \in N$, a server s residing at (i,j) posts at the set

$$P(s) = \{(i,1),...,(i,n)\},$$

and a client c residing at (i,j) queries the set

$$Q(c) = \{(1,j),...,(m,j)\}.$$

(For more examples see [5], which also discusses *truly-distributed, hierarchically-distributed,* etc., name-servers.) We restrict ourselves to methods where the sets $P(s)$ and $Q(c)$ depend on the respective hosts $h(s)$ and $h(c)$ only. It therefore makes more sense to talk about $P(h(s))$ and $Q(h(c))$ instead of $P(s)$ and $Q(c)$. Thus, we define the collection of posting and querying tactics of the set of nodes N, to implement the name-server, as a single *strategy*

$$P,\ Q\ :\ N \rightarrow 2^N,$$

(where 2^N is the set of all subsets of N) for *match-making* in the given network.

1.2. Mutual Exclusion

Another application of the match-making paradigm is distributed *mutual exclusion*. Let the network be as before. In such a distributed system, each network node can issue a mutual exclusion request at an arbitrary time, see e.g. [4]. In order to arbitrate the requests, any pair of two requests must be known to one of the arbitrators. Since these arbitrators must reside in network nodes, any pair of two requests originating from different nodes must reach a common node. Assume that each node i must obtain a permission from each member of a subset $S(i)$ of N before it can proceed to enter its critical section. Then for each pair $(i,j) \in N^2$ we must have $S(i) \cap S(j) \neq \emptyset$ so that the node in the intersection can serve as arbitrator. In [4] the situation is analysed where each node in the network serves as arbitrator equally often, that is, N times. The actual algorithm presented in [4] uses at most $5 \cdot |S(i)|$ messages, where for some K, $|S(i)| = K$ for all i, $i \in N$. (For each set S, $|S|$ will denote the number of elements of S.) It is clear that at least $2K$ messages are required: K messages to query a set $S(i)$, and K answers from every member of $S(i)$ to i. The overhead of $3K$ messages arises from the necessary locking and unlocking protocols to guarantee that no more than one node can simultaneously be in the critical section, to resolve conflicts, and to prevent *deadlock* (i.e., circular waiting among the nodes requesting mutual exclusion) and *starvation* (a node which wants to enter its critical section can be prevented from doing so forever). Here, we may view a *strategy* for distributed mutual exclusion as a mapping

$$S: N \to 2^N$$

and view it as a restricted case of match-making for which the symmetry condition $P(i)=Q(i)$ $(=S(i))$ holds for all $i \in N$.

One way to achieve this symmetry is to let the functions P, Q be as in §1.1, and set $S(i)=P(i) \cup Q(i)$ for all i, $i \in N$. As an example, in the Manhattan network we obtain:

$$S(i,j) = \{(i,1), \ldots, (i,n), (1,j), \ldots, (m,j)\}.$$

More frugal is the example of the *projective plane*. The projective plane $PG(2, k)$ has $n = k^2 + k + 1$ points and equally many lines. Each line consists of $k + 1$ points and $k + 1$ lines pass through each point. Each pair of lines has exactly one point in common (and each pair of points has exactly one line in common). The set of $k+1$ points incident on any of the $k+1$ lines incident on a node i serves as choice for $S(i)$ (see [3,5]).

1.3. Formal Framework

To simplify notation from now on let the set N of network nodes be equal to $\{1,\ldots,n\}$. Without loss of generality, identify s and c with their respective host nodes $h(s)$ and $h(c)$. What is meant, the process of its host, will be clear from the context. For each $s, c \in N$ let $m(s,c)$ be the number of *point-to-point* messages (i.e. messages from a node to any of its direct neighbors) required for the match-making instance of nodes s, c. Then the average number of point-to-point transmissions required for match-making is given by the formula

$$m(n) = \frac{1}{n^2} \sum_{s=1}^{n} \sum_{c=1}^{n} m(s,c). \tag{1a}$$

Since a server s sends messages to all the nodes in $P(s)$ and a client c queries each node in $Q(c)$ the number $m(s,c)$ of point-to-point messages in the match-making instance (s,c) must be at least $|P(s)| + |Q(c)|$. It is exactly the case $m(s,c) = |P(s)| + |Q(c)|$ for which a lower bound is derived in [6]. Another more general situation arises (see [6]) when the average call for a service s by a client c occurs $\alpha(s,c)$-times more often than the average posting of a service available at s. Here one might want to minimize (1a), with $m(s,c) = |P(s)| + \alpha(s,c)\,|Q(c)|$. A similar case arises when in the match-making instance (s,c) the server s is allowed to post $p(s,c)$-many times to the nodes in $P(s)$ and the customer c is allowed to query $q(s,c)$-many times the nodes in $Q(c)$. (This might be necessary in order to increase reliability of the network.) In this case the number $m(s,c)$ of point-to-point messages is equal to $p(s,c)\,|P(s)| + q(s,c)\,|Q(c)|$. The results of the next section are meant to derive lower bounds for these more general situations.

As far as the mathematical aspects of the name-server model are concerned, the operations *posting* and *querying* are entirely symmetric. Thus, one is led to consider *posting*

as a form of *querying*. In general, assume that there are s different forms of querying, say $Q_1,...,Q_s$. Message passing in such an s-dimensional instance $(j_1,...,j_s)$ of the name server mechanism, where $j_r = 1,...,n$, $r = 1,...,s$, is as follows: at the r-th querying the node located at j_r queries a set $Q_r(j_r)$ of nodes. If

$$Q_1(j_1) \cap \cdots \cap Q_s(j_s) \neq \emptyset$$

then any node located at the above intersection will be able to return a message to any of the nodes of $j_1,...,j_s$ stating the address at which the desired service is available.

We can interprete this as a generalization of the name-server: here a client at j_1 queries for a set of services $j_2, \ldots, j_s$, each of a particular nature.

Multimatch-Making

For instance, in the UNIX* system a client may want the TROFF document preparation service, which involves setting up a 'pipe' of preprocessors for a pipelined computation. This may be a pipe like

refer <file>|pic|graph|tbl|eqn|troff

which needs a file, 'refer' preprocessing to assemble references, 'pic' to draw pictures, 'graph' to draw graphs, 'tbl' to make tables, 'eqn' to layout mathematical formulas, and 'troff' to take the final output and assemble an integrated document. Each preprocessor hands the result to the next preprocessor. The client therefore issues requests for many services simultaneously, say for $s-1$ services. We call this *s-match-making*.

Multimutual-Exclusion

Similarly, this approach can be used to investigate *s-mutual exclusion*, i.e. n processes can compete for a resource which can be granted to at most $s-1$ of them for some fixed s, $2 \leq s \leq n$. In this case we want all query sets to be the same set, because of the symmetrical role of the different processors in the algorithm. This is modelled by requiring $Q_1 = \cdots = Q_s$.

This generalization of the critical section problem to the case where at most $s-1$ of processes can be in their critical sections simultaneously was considered in [1]. One way of viewing this problem is to regard it as a resource allocation problem. There are $s-1$ identical copies of of a non-sharable reusable resource, where each process can request at most one copy of that resource. Entry to the critical section corresponds to allocation of a resource copy. Here we can think of each process as having a designated section of code, called the critical section. This code manipulates a resource copy, such that entry of the critical section by a process corresponds to allocation of a resource copy. In [1] the problem is solved through the use of a shared memory which every process can read and write. Deadlock and starvation are avoided. The memory need have only n^2 different values. Essentially, this is a centralized solution. The distributed solution in [4] for

* UNIX is a trademark of AT&T Bell Laboratories.

standard 2-mutual exclusion seems to generalize readily to s-mutual exclusion. The optimal solution there would be something like the projective plane $PG(s,k)$, where the network has $k^s+k^{s-1}+\cdots+1=n$ nodes, each node is incident on k lines, and each line is incident on k nodes. Each s-element subset of lines intersects in precisely one node. Therefore, each query set $S(i)=Q_1(i)=\cdots=Q_s(i)$ of a node i consists of the set of k nodes incident on a line. It does not matter which line we pick, because of the projective plane property that any s lines intersect in one node. The cost in point-to-point messages associated with a particular mutual exclusion instance is therefore at least $sk \approx sn^{1/s}$, which will be shown to be optimal.

However, that is not the topic of the present paper. The problem we address here is that of a lower bound on the number of message passes for each instance of s-mutual exclusion for algorithms of any degree of distributedness within the chosen formal framework.

Formally, the average number of point-to-point messages necessary for match-making (in this more general context) is defined by

$$m(n)=\frac{1}{n^s}\sum_{j_1=1}^{n}\cdots\sum_{j_s=1}^{n}m(j_1,\ldots,j_s), \tag{1b}$$

where $m(j_1,\ldots,j_s)$ is the number of transmissions required for the match-making instance $(j_1,\ldots,j_s)$. (It is an effort to obtain a lower bound for (1b) that motivates the general results of the next section.) In contrast to the *post-query* case, which is best visualized in two dimensions, this more general case is best visualized in s dimensions. (Each axis is marked with the nodes $1,\ldots,n$ and at the vertex $(j_1,\ldots,j_s)$ a point of the intersection $Q_1(j_1)\cap\cdots\cap Q_s(j_s)$ is located.)

The main inequality of the next section is further extended in the last section to obtain lower bounds on the average number of colors occurring across different cross-sections (one for each axis) of a colored (with a finite number of colors) multidimensional body in terms of the total (multidimensional) volume of each color in the whole body. The main motivation for this result is the following. Consider the s-dimensional grid with sides equal to $\{1,\ldots,n\}$, i.e. the cartesian product $\{1,\ldots,n\}^s$; for every s-tuple $(j_1,\ldots,j_s)$ put an element of the intersection $Q_1(j_1)\cap\cdots\cap Q_s(j_s)$ at the vertex $(j_1,\ldots,j_s)$, if the intersection is nonempty, and nothing otherwise. This grid can be considered as a *sufficiently accurate approximation* of a partitioned multidimensional body. Now consider the inequality of theorem 1 below and pass to the limit as the *partitions become finer*.

For general background information on networks the reader should consult [8]. A general discussion of match-making and its relation to mutual exclusion, implementations to different network topologies, as well as a mathematical analysis of the two dimensional case (from which the present research is inspired) is provided in [6]. In addition, the results of the present paper hold for any network topology and for the entire range of

networks, from centralized to distributed.

2. The s-Dimensional Lower Bounds

In this section the main lower bound results are derived. In order to be able to prove the most general results possible it will be necessary to formulate the required concepts with a higher level of abstraction than in the introduction. The motivation however is derived from the discussion in the introduction.

Consider a family $P = \{P_1(j_1),...,P_s(j_s) : j_r = 1,...,n_r,\ r = 1,...,s\}$ of subsets of the set $N = \{1,...,n\}$ of nodes. Let $p_i(j_i) = |P_i(j_i)|$. Let $K_i[P]$ be the set of s-tuples $(j_1,...,j_s)$ such that $i \in P_1(j_1) \cap \cdots \cap P_s(j_s)$ and let $k_i[P] = |K_i[P]|$. (It is clear that if each of these intersections is nonempty then

$$\sum_{i=1}^{n} k_i[P] \geq n_1 \cdots n_s,$$

with $=$ if all $K_i[P]$'s are singleton sets. If all $k_i[P] \in \{0,1\}$ then the left-hand side is $\leq$ the right-hand side.) For the given family P define the *product* (respectively *sum*) $\Pi[P]$ (respectively $A[P]$) corresponding to P by the following formulas:

$$\Pi[P] = \frac{1}{n_1 \cdots n_s} \sum_{j_1=1}^{n_1} \cdots \sum_{j_s=1}^{n_s} p_1(j_1) \cdots p_s(j_s),$$

$$A[P] = \frac{1}{n_1 \cdots n_s} \sum_{j_1=1}^{n_1} \cdots \sum_{j_s=1}^{n_s} [p_1(j_1) + \cdots + p_s(j_s)].$$

Further, for $r = 1,...,s$ put

$$A_r[P] = \frac{1}{n_r} \sum_{j_r=1}^{n_r} p_r(j_r)$$

and notice that

$$\Pi[P] = A_1[P] \cdots A_s[P] \text{ and } A[P] = A_1[P] + \cdots + A_s[P]. \tag{2}$$

The main result of the section is the following

Theorem 1. *For any family P the following inequalities hold:*

$$\Pi[P] \geq \frac{1}{n_1 \cdots n_s} \left(\sum_{i=1}^{n} k_i[P]^{1/s} \right)^s \text{ and } A[P] \geq \frac{s}{(n_1 \cdots n_s)^{1/s}} \left(\sum_{i=1}^{n} k_i[P]^{1/s} \right).$$

Proof: The following inequality, also known as *inequality of the arithmetic and geometric means*, holds for all positive real numbers $a_1, \ldots, a_s$,

$$a_1 + \cdots + a_s \geq s(a_1 \cdots a_s)^{1/s}. \tag{3}$$

In fact, equality holds exactly when all the a_i's are equal (see [3]). For each $r = 1,\ldots,s$ and each $i = 1,\ldots,n$, define the set $H_{r,i}$ consisting of all j_r, $1 \leq j_r \leq n_r$, such that for some $j_1,\ldots,j_{r-1},j_{r+1},\ldots,j_s$,

$$i \in P_1(j_1) \cap \cdots \cap P_{r-1}(j_{r-1}) \cap P_r(j_r) \cap P_{r+1}(j_{r+1}) \cap \cdots \cap P_s(j_s).$$

Also put $h_{r,i} = |H_{r,i}|$. It is now true that for all $i = 1,\ldots,n$,

$$\begin{aligned} h_{1,i} \cdots h_{s,i} &= |H_{1,i} \times \cdots \times H_{s,i}| \\ &\geq |\{(j_1,\ldots,j_s) : i \in P_1(j_1) \cap \cdots \cap P_s(j_s)\}| \\ &= k_i[P]. \end{aligned} \tag{4}$$

Further, for all $r = 1,\ldots,s$,

$$\begin{aligned} \sum_{i=1}^{n} h_{r,i} &\leq \sum_{i=1}^{n} |\{j_r : i \in P_r(j_r)\}| \\ &= \sum_{i=1}^{n} |\{(i,j_r) : i \in P_r(j_r)\}| \\ &= \sum_{j_r=1}^{n_r} |\{(i,j_r) : i \in P_r(j_r)\}| \\ &= \sum_{j_r=1}^{n_r} p_r(j_r) \\ &= n_r \, A_r[P]. \end{aligned} \tag{5}$$

To find the lower bound of $\Pi[P]$ notice that by (5)

$$\begin{aligned} \Pi[P] &= A_1[P] \cdots A_s[P] \\ &\geq \frac{1}{n_1 \cdots n_s} \left(\sum_{i_1=1}^{n} h_{1,i_1} \right) \cdots \left(\sum_{i_s=1}^{n} h_{s,i_s} \right) \\ &= \frac{1}{n_1 \cdots n_s} \sum_{i_1=1}^{n} \cdots \sum_{i_s=1}^{n} h_{1,i_1} \cdots h_{s,i_s} \\ &= \frac{1}{n_1 \cdots n_s} \sum_{i_1=1}^{n} \cdots \sum_{i_s=1}^{n} S(i_1,\ldots,i_s), \end{aligned}$$

where $S(i_1,\ldots,i_s) = h_{1,i_1} \cdots h_{s,i_s}$. By cyclically rotating the indices $i_1,\ldots,i_s$ one obtains the following s-many summands:

$$S(i_1,\ldots,i_{s-1},i_s),\ S(i_2,\ldots,i_s,i_1),\ \ldots,\ S(i_{s-1},\ldots,i_{s-3},i_{s-2}),\ S(i_s,\ldots,i_{s-2},i_{s-1}).$$

Using inequalities (3) and (4) it is easy to see that the sum of the above s-many summands must be at least

$$s(k_{i_1}[P] \cdots k_{i_s}[P])^{1/s}.$$

Adding the above s-many summands with respect to $i_1,\ldots,i_s$ and taking into account the last inequality one easily obtains that

$$\sum_{i_1=1}^{n} \cdots \sum_{i_s=1}^{n} h_{1,i_1} \cdots h_{s,i_s} \geq \sum_{i_1=1}^{n} \cdots \sum_{i_s=1}^{n} (k_{i_1}[P] \cdots k_{i_s}[P])^{1/s} = \left(\sum_{i=1}^{n} k_i[P]^{1/s} \right)^s .$$

This completes the proof of the lower bound of $\Pi[P]$. The lower bound on $A[P]$ is an immediate consequence of equations (2), inequality (3) and the lower bound for $\Pi[P]$. This completes the proof of the theorem •

In particular, both propositions 1 and 2 of [6] are immediate consequences of theorem 1. The reader familiar with [6] will undoubtedly notice that the proof of theorem 1 is essentially a generalization of proposition 1 of [6] to rectangles (possibly with holes) and dimensions $s \geq 2$.

Example 1. *(Multidimensinal Cube Network)* Let the number of nodes be $n = 2^d$ and suppose that the number s of queries is a divisor of d. Addresses of nodes consist of d bits, like $u_1 u_2 \cdots u_d$. Nodes are connected by an edge exactly when they differ by a single bit. For each $r = 1,...,s$, let $Q_r(u_1 \cdots u_d)$ be the set

$$\{x_1 \cdots x_{(r-1)d/s}\, u_{(r-1)d/s+1} \cdots u_{rd/s}\, x_{rd/s+1} \cdots x_d : x_i \in \{0,1\}\}.$$

Clearly, each of the above sets has size $2^{(s-1)d/s}$ and $|k_i[P]| = 2^{(s-1)d} = n^{s-1}$. Thus, with $m(j_1,...,j_s) = |P_1(j_1)| + \cdots + |P_s(j_s)|$, one easily obtains that $m(n) \geq sn^{(s-1)/s}$, i.e. the average number of point-to-point message transmissions is at least $sn^{(s-1)/s}$.

Example 2. It is easy to see that for any family P, $\Pi[P] \leq n^s$, $A[P] \leq sn$. If in addition, $P_r(j_r) = \{1,...,n\}$, for all $j_r = 1,...,n_r$, $r = 1,...,s$, then the lower bounds of theorem 1 are identical to the upper bounds given above. Consequently, the lower bounds of theorem 1 are optimal (see also the remark below).

Example 3. Let the assumptions be as in Theorem 1. Let $n_1 = n_2 = \cdots = n_s$. Then the average number of colors in an axis parallel cross-section of P equals

$$\Pi(P)^{1/s} \ (= \frac{A[P]}{s})$$

and Theorem 1 provides a lower bound on this number.

Remark. In the statement of theorem 1, the quantity $A[P]$ equals the right-hand side of the inequality in which it occurs, exactly when $A_1[P] = \cdots = A_s[P]$. (This is an immediate consequence of the inequality on arithmetic and geometric means.) In particular, the optimal name-servers are the ones for which the average number of point-to-point transmissions are equally balanced in all directions. This remark also applies to theorem 5, below.

3. Applications to Distributed Match-Making

As an application of theorem 1 one can determine lower bounds for sums of the form

$$S[P] = \frac{1}{n_1 \cdots n_s} \sum_{j_1=1}^{n_1} \cdots \sum_{j_s=1}^{n_s} [p_1(j_1,\ldots,j_s)p_1(j_s) + \cdots + p_s(j_1,\ldots,j_s)p_s(j_s)], \quad (6)$$

where $p_r(j_1,\ldots,j_s)$ is a positive integer for each $r = 1,\ldots,s$ and each $j_r = 1,\ldots,n_r$. (These are the types of sums for which a lower bound was promised in the introduction.) To determine the lower bound on $S[P]$ put

$$N_{r,j_r} = \sum_{j_1=1}^{n_1} \cdots \sum_{j_{r-1}=1}^{n_{r-1}} \sum_{j_{r+1}=1}^{n_{r+1}} \cdots \sum_{j_s=1}^{n_s} p_r(j_1,\ldots,j_s) \text{ and } N_r = \sum_{j_r=1}^{n_r} N_{r,j_r},$$

for each $r = 1,\ldots,s$. It is now clear that

$$S[P] = \frac{1}{n_1 \cdots n_s} \left(\sum_{j_1=1}^{n_1} N_{1,j_1}p_1(j_1) + \cdots + \sum_{j_s=1}^{n_s} N_{s,j_s}p_s(j_s) \right).$$

Consider a new family $Q = \{Q_r(t_r) : r = 1,\ldots,s,\ t_r = 1,\ldots,N_r\}$, where $Q_r(t_r) = P_r(j_r)$ for the unique j_r such that

$$[1 \leq t_r \leq N_{r,1} \text{ and } j_r = 1]$$

or

$$[N_{r,1} + \cdots + N_{r,j_r-1} < t_r \leq N_{r,1} + \cdots + N_{r,j_r} \text{ and } n_r \geq j_r > 1].$$

As before, let $q_r(t_r) = |Q_r(t_r)|$. It is now clear from inequality (3) and theorem 1 that

$$\begin{aligned} S[P] &= \frac{1}{n_1 \cdots n_s} \left(\sum_{t_1=1}^{N_1} q_1(t_1) + \cdots + \sum_{t_s=1}^{N_s} q_s(t_s) \right) \\ &\geq \frac{s}{n_1 \cdots n_s} \left(\left(\sum_{t_1=1}^{N_1} q_1(t_1) \right) \cdots \left(\sum_{t_s=1}^{N_s} q_s(t_s) \right) \right)^{1/s} \\ &= \frac{s}{n_1 \cdots n_s} (N_1 \cdots N_s)^{1/s} \Pi[Q]^{1/s} \\ &= \frac{s}{n_1 \cdots n_s} \sum_{i=1}^{n} k_i[Q]^{1/s}. \end{aligned}$$

It remains to compare the quantities $k_i[P]$, $k_i[Q]$. This can be done by comparing the sizes of the sets $K_i[P]$, $K_i[Q]$. Indeed, by definition of the family Q to each $(j_1,\ldots,j_s) \in K_i[P]$ there correspond at least $N_{1,j_1} \cdots N_{s,j_s}$-many s-tuples $(t_1,\ldots,t_s)$ of the set $K_i[Q]$. Hence, the following theorem has been proved.

Theorem 2. *For any family P the following inequality holds:*

$$S[P] \geq \frac{s}{n_1 \cdots n_s} \sum_{i=1}^{n} \left(\sum \{ N_{1,j_1} N_{2,j_2} \cdots N_{s,j_s} : (j_1,\ldots,j_s) \in K_i[P] \} \right)^{1/s} \bullet$$

Now it is easy to obtain lower bounds for the average number of point-to-point messages for distributed match-making (this will answer open questions posed in [6]). For example, one can handle the case

$$m(i,j) = p(i,j) \mid P(i) \mid + q(i,j) \mid Q(j) \mid$$

mentioned in the introduction by using theorem 2. More generally, one can study the case

$$m(j_1,\ldots,j_s) = p_1(j_1,\ldots,j_s) \mid P_1(j_1) \mid + \cdots + p_s(j_1,\ldots,j_s) \mid P_s(j_s) \mid .$$

Thus, for example, if each $p_r(j_1,\ldots,j_s) = \lambda_r$ then the average amount $m(n)$ of point-to-point messages, as defined by equation (1b), must satisfy the inequality of the following

Corollary 3.

$$m(n) \geq \frac{s(\lambda_1 \cdots \lambda_s)^{1/s}}{n} \sum_{i=1}^{n} k_i^{1/s},$$

where $k_i = \mid \{(j_1,\ldots,j_s) : i \in P_1(j_1) \cap \cdots \cap P_s(j_s)\} \mid \bullet$

Let $n_1 = n_2 = \cdots = n_s = n$. For really distributed s-mutual exclusion we require all $p_i[j_i]$'s to be equal, say K, and therefore $A[P] = sK$ $(1 \leq i, j_i \leq n)$. Now expressing $A[P]$ in the $k_i[P]$'s which measure the distributedness of the algorithm:

$$A[P] = sK \geq \frac{s}{n} \sum_{i=1}^{n} k_i[P]^{1/s}$$

Setting all $k_i[P]$'s equal, gives therefore: $K \geq n^{(s-1)/s}$. The resources in global storage space used must be essentially $n^s sK \geq sn^{s+(s-1)/s}$ but locally only sK is used.

4. Application to Colored Multidimensional Bodies

It is natural to think of the result of theorem 1 as providing a lower bound on the average number of colors occurring across the different cross-sections of a *colored multidimensional grid* in terms of the total number of occurrences of each color in the *whole grid.* Now, it is desired to extend this result to *finitely colored, continuous, multidimensional bodies.* One way to do this, is to partition the given multidimensional body into *infinitely small* multidimensional cubes, apply theorem 1 to the resulting grid and pass to the limit, as the *size* of the *members of the partition* becomes infinitely small. Although this argument works, it suffers from two drawbacks. First, it causes notational

complications and second, the class of bodies for which the above limits exist is much smaller than the class of *Lebesgue measurable sets*. Thus, in order to avoid both the complications and limitations which arise from the possible nonexistence of such limits it will be necessary to use the notion of *Lebesgue measure* (see [2]).

For any *Lebesgue measurable* set $S \subseteq R^s$ let $m(S)$ denote its *Lebesgue measure*. (In the sequel, the same symbol will be used for r-dimensional Lebesgue measure on R^r, i.e. Lebesgue measure on sets of r-tuples of real numbers, for each $r = 1,...,s$; however, this will cause no confusion because it will be clear from the context which measure is meant in each case. In addition, all integrations considered below are with respect to Lebesgue measure.)

Let $B \subseteq R^s$ be a Lebesgue measurable set. For each $r = 1,...,s$ let B_r be the projection of B onto the x_r-axis, i.e. B_r is the set of u_r such that for some $u_1,...,u_{r-1},u_{r+1},...,u_s$, $(u_1,...,u_s) \in B$. For each r and each u_r let $B_r(u_r)$ be the u_r cross-section of B, i.e..

$$B_r(u_r) = \{(u_1,...,u_{r-1},u_{r+1},...,u_s) : (u_1,...,u_s) \in B\}.$$

(Thus, each $B_r \subseteq R$ and each $B_r(u_r) \subseteq R^{s-1}$.)

Suppose that B is colored with n colors, say $1,....,n$. For $r = 1,...,s$ let $P_r(u_r) = \{i :$ color i occurs in $B_r(u_r)\}$ and put $p_r(u_r) = |P_r(u_r)|$. Further, it is assumed that each set

$$K_i[B] = \{(u_1,...,u_s) \in B : (u_1,...,u_s) \textit{ is colored with color } i\}$$

is Lebesgue measurable, where $i = 1,...,n$. Put $k_i[B] = m(K_i[B])$. (As in section 2 it is clear that $\sum_{i=1}^{n} k_i[B] = m(B)$.) Since,

$$B_1(u_1) \cap \cdots \cap B_s(u_s) = \{(u_1,...,u_s)\},$$

it is evident that $P_1(u_1) \cap \cdots \cap P_s(u_s) = \{i\}$, where i is the (unique) color of vertex $(u_1,...,u_s)$.

For each $r = 1,...,s$ define

$$A_r[B] = \frac{1}{m(B_r)} \int_{B_r} p_r(x_r)\,dx_r .$$

Further, as in section 2, put

$$\Pi[B] = \frac{1}{m(B_1) \cdots m(B_s)} \int_{B_1} \cdots \int_{B_s} p_1(x_1) \cdots p_s(x_s)\,dx_1 \ldots dx_s ,$$

$$A[B] = \frac{1}{m(B_1) \cdots m(B_s)} \int_{B_1} \cdots \int_{B_s} [p_1(x_1)) + \cdots + p_s(x_s)]\,dx_1 \ldots dx_s$$

and notice that

$$\Pi[B] = A_1[B] \cdots A_s[B],\ A[B] = A_1[B] + \cdots + A_s[B].$$

In addition, let $B_{r,i}$ be the set of all u_r such that for some $u_1,\ldots,u_{r-1},u_{r+1},\ldots,u_s$, the s-tuple $(u_1,\ldots,u_{r-1},u_r,u_{r+1},\ldots,u_s)$ is colored with color i. (At this point the reader should be aware of the apparent similarities between the sets $H_{r,i}$, defined in the proof of theorem 1, and the sets $B_{r,i}$ defined here.) As before it can be shown that for all $i = 1,\ldots,n$, $r = 1,\ldots,s$,

$$m(B_{1,i}) \cdots m(B_{s,i}) \geq k_i[B], \tag{7}$$

$$\sum_{i=1}^{n} m(B_{r,i}) \leq m(B_r) \cdot A_r[B]. \tag{8}$$

The proof of (7) is trivial; the proof of (8) is similar to the proof of (5) but it requires the following lemma, which is proved by induction on n.

Lemma 4. *For any Lebesgue measurable sets $S_1,\ldots,S_n$,*

$$\sum_{i=1}^{n} m(S_i) \leq \sum_{k=1}^{n} k \cdot m(\{u : u \text{ belongs to exactly } k\text{-many } S_i\text{'s}\}) \bullet$$

(The reader may find it convenient to convince himself of the validity of lemma 4 by drawing a picture for the case $s = 2$, $n = 3$.) Now inequality (8) is an immediate consequence of the lemma. Indeed,

$$\begin{aligned}\sum_{i=1}^{n} m(B_{r,i}) &\leq \sum_{i=1}^{n} m(\{u_r \in B_r : i \in P_r(u_r)\}) \\ &\leq \sum_{k=1}^{n} k \cdot m(\{u_r \in B_r : |P_r(u_r)| = k\}) \\ &= \int_{B_r} p_r(u_r)\,du_r .\end{aligned}$$

Finally, using inequalities (7), (8) and arguing as in the last part of the proof of theorem 1 it is easy to obtain the following

Theorem 5. *The following inequalities hold for any Lebesgue measurable set $B \subseteq R^s$, and any partition (or coloring) of B into n Lebesgue measurable subsets:*

$$\Pi[B] \geq \frac{1}{m(B_1) \cdots m(B_s)} \left(\sum_{i=1}^{n} (k_i[B])^{1/s} \right)^s ,$$

$$A[B] \geq \frac{s}{(m(B_1) \cdots m(B_s))^{1/s}} \left(\sum_{i=1}^{n} (k_i[B])^{1/s} \right) \bullet$$

The following example might help illustrate the concepts involved.

Example 4. Let B be the open disc with center at (0,0) and radius 3 units. Color B with the three colors 1,2,3. For each $i = 1,2,3$ let $K_i[B]$ be the set of all pairs (x,y) such

that

$i-1 \leq \sqrt{x^2+y^2} < i$. Using the notation of theorem 5, it is easy to show that $k_1[B] = \pi$, $k_2[B] = 3\pi$, $k_3[B] = 5\pi$, $m(B_1) = m(B_2) = 6$. It follows from the second inequality of theorem 5 that $A[B] \geq 2.7468$.

Since it is immediately clear that Theorem 5 holds even when we rotate or translate the axis, we have:

Example 5. Let $m(B_1)=m(B_2)=\cdots=m(B_s)$ and let the assumptions be as in Theorem 5. Then the average number of colors in a cross-section of B is equal to

$$\Pi(B)^{1/s} \ (= \frac{A[B]}{s})$$

Therefore, the average number of colors in a cross-section of the disc of example 4 is ≥ 1.3734.

The reader familiar with classical measure theory and product measures (see [2]) should have no difficulty extending this last theorem to arbitrary, positive, countably-additive measures μ. One simply replaces m with μ in the proof of theorem 5. The resulting theorem generalizes both theorems 1 and 5 (theorem 1 corresponds to $\mu =$ *counting measure* and theorem 5 corresponds to $\mu =$ *Lebesgue measure*). Details are left to the reader.

References

[1] Fischer, M.J., Lynch, N.A., Burns, J.E., and Borodin, A., *Distributed FIFO allocation of identical resources using small shared space*, Massachusetts Institute of Technology, Cambridge, Mass., Report MIT/LCS/TM-290, October, 1985.

[2] Halmos, P. R., *Measure Theory*, Springer-Verlag, 1974.

[3] Hardy, G. H., Littlewood, J. E. and Polyá, G., *Inequalities*, Cambridge University Press, 1934.

[4] Maekawa, M., *A $\sqrt{N}$ Algorithm for Mutual Exclusion in Decentralized Systems*, ACM Transactions on Computer Systems **3** (1985), pp. 145-159.

[5] Mullender, S. J., *Principles of Distributed Operating System Design* , Ph.D. Thesis, Vrije Universiteit te Amsterdam, 1985.

[6] Mullender, S. J. and Vitányi, P. M. B., *Distributed Match-Making for Processes in Computer Networks*, Proceedings of the 4th annual ACM Symposium on Principles of Distributed Computing, 1985, pp. 261-271.

[7] Powell, M.L. and Miller, B. P., *Process Migration in DEMOS/MP*, Proceedings of the 9th ACM Symposium on Operating Systems Principles, 1983, pp. 110-119.

[8] Tanenbaum, A. S., *Computer Networks*, Prentice-Hall, 1981.

[9] Mullender, S.J. and A.S. Tanenbaum, *The Design of a Capability Based Distributed Operating System,* The Computer Journal, 29(1986) pp. xxx-xxx

OFFICIAL DISTRIBUTION LIST

1985

Director
Information Processing Techniques Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

2 Copies

Office of Naval Research
800 North Quincy Street
Arlington, VA 22217
Attn: Dr. R. Grafton, Code 433

2 Copies

Director, Code 2627
Naval Research Laboratory
Washington, DC 20375

6 Copies

Defense Technical Information Center
Cameron Station
Alexandria, VA 22314

12 Copies

National Science Foundation
Office of Computing Activities
1800 G. Street, N.W.
Washington, DC 20550
Attn: Program Director

2 Copies

Dr. E.B. Royce, Code 38
Head, Research Department
Naval Weapons Center
China Lake, CA 93555

1 Copy

Dr. G. Hopper, USNR
NAVDAC-OOH
Department of the Navy
Washington, DC 20374

1 Copy

END

11-86

DTIC