Distributed FIFO Allocation of Identical Resources Using Small Shared Space

MICHAEL J. FISCHER
Yale University
NANCY A. LYNCH
Massachusetts Institute of Technology
JAMES E. BURNS
Georgia Institute of Technology
and
ALLAN BORODIN
University of Toronto

We present a simple and efficient algorithm for the FIFO allocation of k identical resources among asynchronous processes that communicate via shared memory. The algorithm simulates a shared queue but uses exponentially fewer shared memory values, resulting in practical savings of time and space as well as program complexity. The algorithm is robust against process failure through unannounced stopping, making it attractive also for use in an environment of processes of widely differing speeds. In addition to its practical advantages, we show that for fixed k, the shared space complexity of the algorithm as a function of the number N of processes is optimal to within a constant factor.

Categories and Subject Descriptors: C.2.2 [**Computer-Communication Networks**]: Network Protocols—*protocol architecture*; C.2.4 [**Computer-Communication Networks**]: Distributed Systems—*distributed applications, network operating systems*; C.4 [**Computer Systems Organization**]: Performance of Systems—*reliability, availability, and servicability*; D.4.1 [**Operating Systems**]: Process Management—*concurrency, mutual exclusion, synchronization*; D.4.5 [**Operating Systems**]: Reliability—*fault-tolerance*; F.1.2 [**Computation by Abstract Devices**]: Modes of Computation—*parallelism*

General Terms: Algorithms, Performance, Reliability, Theory

Additional Key Words and Phrases: Asynchronous system, distributed computing, FIFO, lower bound, queue, resource allocation, shared memory, space complexity

This work was supported in part by the Office of Naval Research under contract N00014-82-K-0154; by the U.S. Army Research Office under contract DAAG29-79-C-0155; and by the National Science Foundation under grants MCS77-02474, MCS77-15628, MCS78-01689, MCS-8116678, and DCR-8405478. N. A. Lynch's work was supported by NSF grant CCR-8611442, DARPA N00014-83-K-0125, and ONR N00014-85-K-0168.

Authors' current addresses: M. J. Fischer, Department of Computer Science, Yale University, 10 Hillhouse Avenue, P.O. Box 2158, Yale Station, New Haven, CT 06520-2158; N. A. Lynch, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139; J. E. Burns, School of Information and Computer Science, Georgia Institute of Technology, Atlanta, GA 30332-0280; A. Borodin, Department of Computer Science, Sanford Fleming Building, 10 Kings College Circle, Room 2303B, Toronto, Ontario M5S 1A4.

1. INTRODUCTION

1.1 Critical Section Problem

The critical section problem has been widely studied for its illustrative value in problems of synchronization as well as for its practical application to real concurrent systems [1–15, 18–22]. The problem is to devise protocols for each of several communicating asynchronous parallel processes to control access to a designated section of code called the *critical section.* Such code might manipulate a common resource, in which case access to the critical section corresponds to allocation of the resource. In the simple case of a single nonsharable, reusable resource (such as a line printer or a tape drive), the two basic properties desired of the access policy are *mutual exclusion* and *impossibility of deadlock.* Mutual exclusion means that two processes can never simultaneously be executing their critical sections. Deadlock is a situation in which one or more processes are attempting to enter or leave their critical sections, but none of them ever succeeds. Finding appropriate protocols to insure these two properties is the *critical section problem.*

Two protocols comprise a solution. The *trying protocol* is the section of code that a process executes before being admitted to its critical section, and the *exit protocol* is the code to be run when the process leaves its critical section and returns to the remainder of its code, called the *remainder section.* Equivalently, the trying protocol allocates the resource corresponding to the critical section and the exit protocol returns it to the system.

1.2 Multiple Resources

In this paper, we generalize the critical section problem to the case where some number $k \geq 1$ of processes (but not more) are permitted to be simultaneously in their critical sections. Regarded as a resource-allocation problem, we consider k identical copies of a nonsharable reusable resource, where each process can request at most one copy of that resource. Again entry to the critical section corresponds to allocation of a resource copy, but we ignore questions of just how the individual copies of the resource are managed.

The exclusion property of the k-critical section problem, that at most k processes are ever simultaneously in their critical sections, we call *k-exclusion.* To avoid degenerate solutions, we must also formalize the notion that "it should be possible for as many as k processes to be simultaneously in their critical sections." We interpret this to mean, roughly, that if fewer than k processes are in their critical sections, then it is possible for another process to enter its critical section, even though no process leaves its critical section in the meantime. We call this property "avoiding k-deadlock." Precise definitions of these properties are deferred until Section 3, after the algorithms have been presented.

1.3 Naive Solutions

We first consider two obvious solutions for handling multiple resources and examine their shortcomings.

1.3.1 *Generalized Semaphore.* A trivial generalization of a binary semaphore yields a system exhibiting k-exclusion and no k-deadlock. Assume a shared variable, COUNT, which at any time contains the correct count of the number of processes currently in their critical sections. A process wanting to enter its critical section performs an atomic action on COUNT which, in one indivisible step, reads the value of COUNT, increments it if it was less than k, and stores back the result. The process then proceeds to its critical section if it saw COUNT less than k, and it loops back and repeats the test otherwise (busy-waiting). A process leaving its critical section simply decrements COUNT.

This algorithm imposes no fairness criteria on the order in which processes enter their critical sections, and in fact it is possible that an individual process will always find the critical section "full" (i.e., COUNT $= k$) whenever it happens to examine COUNT and therefore will be "locked out" of its critical section forever. It would be much more interesting to have algorithms that exhibited fair behavior, guaranteeing, for example, that no process is locked out or that access to the critical section is FIFO with respect to the time when processes first begin their trying protocols.

1.3.2 *Bank Algorithm.* Rather than devise new algorithms for the k-critical section problem with stronger fairness conditions, an obvious approach is to try to reduce the k-critical section problem to the 1-critical section problem and then apply known solutions to the latter problem, e.g., [2–5], and [11]. Such a hybrid solution is commonly used in banks for scheduling people waiting for a teller. People entering the bank line up in a single queue. When one or more tellers become available, the person at the head of the queue goes to any free teller.

To see the reduction that is illustrated by this simple example, think of the position at the head of the queue as a "resource." Only one person has this resource at a time, and the queue itself serves to allocate that resource in first-in-first-out (FIFO) order. Only the person holding the head-of-queue resource is permitted to go to a teller, so the order of service by a teller is "essentially" FIFO, modulo possible delay between leaving the head of the queue and arriving at a teller.[1] Such a reduction is possible given any 1-critical section solution, and the number of values of shared memory increases by only a factor of $(k + 1)$.

The bank algorithm has a rather subtle defect that becomes apparent when several tellers become simultaneously free. If $k \geq 2$ tellers are free, one would like the first k people in line to all move "simultaneously" to a teller, yet the algorithm requires them to file past the head of the queue one at a time. If the person at the front of the line is slow, the $k - 1$ people behind him are forced to wait unnecessarily. In fact, if the person at the front of the line "fails," then the people behind him wait forever and the system stops functioning. In this case, one failure can tie up all of the system's resources!

1.4 Additional Robustness and Fairness Requirements

The previous section leads us to impose additional requirements to control the degradation of processing in the event that a limited number of processes fail during the execution of their protocols and to provide for fairness of access to the critical section.

[1] By running, a person might actually arrive at a teller before another who was ahead of him in the queue. Nevertheless we consider this to be a reasonable approximation of what people mean by FIFO since once one arrives at the front of the queue, one no longer has to wait for others.

Our notion of "failure" is quite different from the "shutdown" considered in [21] and [22]. Unlike a process that shuts down, a failed process does not announce to the world that it has failed. Rather, we say a process *fails* if there is a time after which it executes no more steps of its program. To distinguish a failed process from a correct one that is merely running very slowly, one must look infinitely far into the future and determine that it never takes another step. Thus, other processes have no way of distinguishing a failed process from a correct one in a finite amount of time. (In particular, timeouts won't work since we make no assumptions about the relative speeds of processes.)

Our interest in this kind of failure stems partly from the practical problems of building fault-tolerant distributed systems and partly from the desire to understand the dependencies among processes competing for entry to their critical sections. Each instance of one process waiting for another indicates a lack of concurrency in the whole solution. Taken together, these dependencies tend to cause the whole system to run at the speed of the slowest process. Algorithms that continue to operate correctly even when a limited number of processes fail cannot exhibit such simple dependencies. For example, if process A waits for process B to take some action and process B were to fail, then process A would wait forever and make no further progress toward its goal. Hence B's failure would cause the whole system to fail by locking out A. Insisting that algorithms be robust in the face of a certain amount of failure gives us a formal way of studying degrees of concurrency, which in turn have implications for the running time of the system.

At first sight, the concepts of robustness and fairness, say FIFO ordering, appear to be contradictory. Robustness says, for example, that if one process fails in its trying protocol, the system must continue to function. In particular, other processes that enter their trying protocols after the failed one will necessarily enter their critical sections ahead of it. Since the apparently failed process might actually be correct but slow, robustness implies a violation of the usual definition of FIFO ordering.

The problem is circumvented by defining the fairness conditions not in terms of the order in which processes *enter* their critical sections but rather by the order in which they become "enabled" to enter their critical sections. By *enabled,* we mean that a process no longer needs to wait for action by any other process before it can go into its critical section, nor can the actions or inactions of other processes prevent it from doing so. Intuitively, when a process becomes enabled, a copy of the resource is reserved for it, and actions of other processes are no longer needed in order for the given process to complete its trying protocol. The key distinction between enabling and actual entry to the critical section is that a process might become enabled passively by the action of some other process changing shared memory, whereas entry to the critical section can take place only by a positive action of the given process.

1.5 Colored Ticket Algorithm

In this paper, we describe an algorithm, the *Colored Ticket Algorithm,* for solving the k-critical section problem. It is robust, enables processes in FIFO order, and uses $O(N^2)$ values of shared memory for fixed k. The algorithm can be thought of as a distributed implementation of a queue, for it simulates the behavior that would be achieved by explicitly storing the entire queue of waiting processes in shared memory, but it uses far less shared space and is fast and simple to implement. We also show for fixed k that our algorithm is essentially optimal in terms of the amount of shared memory used by giving an $\Omega(N^2)$ lower bound on the number of distinct shared memory values for any robust algorithm that so simulates a queue.

If one weakens the robustness conditions to permit lockout to occur in case more than a prespecified number of processes fail, then more space-economical solutions are possible [9]. However these solutions lack the elegance and simplicity of the Colored Ticket Algorithm as well as its time efficiency. If one ignores robustness altogether, then $O(N)$ values suffice [17].

1.6 Synopsis

The main technical content of the paper is contained in the next four sections. The Colored Ticket Algorithm is presented in Section 2 by a series of refinements, starting with a straightforward queue algorithm (Figure 1) and ending with the full algorithm (Figure 7). The correctness of the resulting algorithm follows from the correctness of each of the refinement steps.

The remainder of the paper develops the formal machinery needed to substantiate our space complexity claims for the Colored Ticket Algorithm. Section 3 presents a formal model of computation and precise definitions of properties that characterize the k-critical section problem. Section 4 describes how to translate the Colored Ticket Algorithm to a process in the formal model and sketches how to prove that it solves the k-critical section problem in small shared space. Section 5 contains the lower bound proof that shows when k is fixed, the Colored Ticket Algorithm is space-optimal (to within a constant factor depending exponentially on k) among all algorithms that solve the strong form of the k-critical section problem given in this paper.

2. k-CRITICAL SECTION ALGORITHMS

In this section, we present the Colored Ticket Algorithm, using a process of successive refinement. Specifically, we present four algorithms, each a refinement of the preceding, the last of which is the Colored Ticket Algorithm. The algorithms run in an environment consisting of N processes, each with its own private memory, and a common shared memory through which the processes communicate. Access to the shared memory is via atomic actions that allow a process, in one indivisible step, to read the entire contents of shared memory and modify it in an arbitrary way, depending in general on the data just read.

We specify our algorithms in a Pascal-like language augmented with two new statements for specifying atomic actions, **start** and **finish**. Statements executed dynamically after **start** and before the next **finish** comprise a single *atomic action*. While it is possible in this language to write atomic actions of unbounded size (for example, by executing a loop between **start** and **finish**), the atomic actions we actually use are all bounded, a fact that is important for the implementation which we give in Section 3.

```
repeat forever
    start;
    ENQUEUE($i$);
    wait until $i$ is in one of the first $k$ positions of QUEUE;
    finish;

    { Critical Section }

    start;
    REMOVE($i$);
    finish;

    { Remainder Section }

end repeat.
```

Fig. 1. Queue algorithm (code for process i).

In order for our algorithms to have the desired correctness and robustness properties, we make two assumptions about the implementation of atomic actions:

(1) A process crash in the middle of an atomic action does not cause the system to hang and leaves the shared memory as it was before beginning the action.
(2) The system never aborts atomic actions. (Alternatively, an atomic action that is retried repeatedly will eventually succeed.)

While these assumptions are difficult to implement exactly, they can be approximated in real systems, so we believe our algorithms will be useful in practice.

As a convenience, we use the construct "**wait until** C" as an abbreviation for "**while not** C **do** [**finish**; **start**]". Thus every execution of the wait loop ends one atomic action and begins another.

2.1 The Queue Algorithm

We first describe a simple but inefficient solution to the k-critical section problem. This basic algorithm, the Queue Algorithm, stores the entire queue of waiting and critical processes in the shared variable. A process in any of the first k positions of the queue is permitted to enter its critical section. This algorithm requires no communication among processes other than that provided by the queue itself, and, in fact, each process need only change shared memory at the moments of entry to the trying protocol and remainder section.

In the code given in Figure 1, the shared memory contains a single queue which admits two operations. ENQUEUE places an element at the rear of the queue, and REMOVE deletes a particular element from the queue, regardless of where it occurs. Initially the queue is empty.

Note that many atomic actions might be executed before the process reaches its critical section since each execution of the wait loop ends one atomic action and begins another. However, only the first of these actually updates shared memory; the others are all "read-only."

In this very simple algorithm, a process P_i becomes *enabled* to enter its critical section when i first appears in one of the first k positions of the queue. Enabling can occur either when i enters the queue, if fewer than k process numbers are in the queue at that time, or when REMOVE(j) is executed for some j that is ahead of i in the queue. Once enabled, P_i remains so until REMOVE(i) is executed, for any new process numbers enter the queue behind i.

It should be clear that this algorithm achieves the robustness and fairness conditions described in the introduction. Processes become enabled in the same order as they entered the queue; and once enabled, a process remains enabled until it finishes its critical section. Processes cannot be locked out of their critical sections unless all of the first k positions of the queue are filled with failed processes.

2.2 Ticket Systems

While the Queue Algorithm satisfies all the correctness properties we want, keeping the queue in shared memory requires too much space to make the algorithm very interesting. Our goal is to find an algorithm equivalent to the Queue Algorithm that keeps a lot less information in the shared variable. In other words, we wish to devise a space-efficient "distributed simulation" of the Queue Algorithm.

All of our remaining algorithms are *ticket systems,* modeled after the ticket systems often used in bakeries. A process wishing to enter its critical section takes the next available *ticket* from an ordered sequence of tickets and then waits until its ticket becomes *valid,* at which point it is enabled to enter its critical section and proceeds to do so on its next step. When it leaves its critical section, it discards its ticket and *validates* the next invalid ticket in order, thereby allowing the next process in line to proceed. (In case no process is currently waiting, the next ticket is nevertheless validated, and when a process eventually takes that validated ticket, it will proceed directly to its critical section.) Once a ticket becomes valid, it remains valid until discarded. Tickets are validated in the same order as they are issued, and at any time, exactly k (nondiscarded) tickets are valid, some of which may not have yet been issued.

Since every process in its critical section holds a valid ticket, k-exclusion is satisfied. Since tickets are validated in the order in which they are issued, processes are enabled in FIFO order, so the algorithm satisfies our fairness condition. Robustness follows since a process does not modify shared memory from the time it enters its trying protocol until the time it returns to its remainder section; hence, whether or not it is alive in the meantime has no effect on the rest of the system.

Any such ticket algorithm simulates the Queue Algorithm in the sense that if the natural correspondence is made between atomic actions of the two algorithms and those actions are run in the same order, then processes enter and leave their critical and remainder sections in exactly the same order in both. Indeed, the simulated queue of the Queue Algorithm can be obtained by arranging the processes holding tickets in increasing order of their tickets. Issuing a ticket corresponds to adding a process to the end of the queue, and discarding a ticket

```
local variable TICKET;

repeat forever

    start;
    TICKET := TAKE_NEXT_TICKET;
    wait until IS_VALID(TICKET);
    finish;

    { Critical Section }

    start;
    VALIDATE_NEXT_TICKET(TICKET);
    finish;

    { Remainder Section };

end repeat.
```

Fig. 2. Basic ticket algorithm (code for process *i*).

together with validating the next (invalid) ticket corresponds to removing a process from the queue. The *k* valid tickets always correspond to the first *k* positions of the queue.

Code to implement this basic paradigm is shown in Figure 2.

Function TAKE_NEXT_TICKET *issues* the next available ticket and returns it to the calling program. Function IS_VALID(T) returns a Boolean value telling whether or not the ticket T is valid. Procedure VALIDATE_NEXT_TICKET(T) discards the ticket T and updates shared memory so as to cause the next invalid ticket in sequence to become valid. In order to fully specify a ticket algorithm, one must specify these three subroutines. (Initial conditions must guarantee that TAKE_NEXT_TICKET will be ready to issue exactly *k* valid tickets.)

2.3 The Numbered Ticket Algorithm

The first ticket system we present, the *Numbered Ticket Algorithm,* uses an infinite number of values and hence requires an unbounded amount of shared memory for its implementation. The Colored Ticket algorithm, which uses only finite shared memory, is then described as two further modifications of this algorithm.

In the Numbered Ticket Algorithm, tickets are natural numbers in their usual order. The algorithm maintains two variables in shared memory. ISSUE holds the most recently issued ticket, and VALID holds the most recently validated ticket. Initially ISSUE = 0 and VALID = k. An entering process takes a ticket by incrementing ISSUE and using the variable's new value as its ticket number. Ticket number t is valid whenever VALID $\geq t$; hence, any process can determine by looking in shared memory whether or not its ticket is valid. A process returning to its remainder section discards its ticket and increments VALID.

```
global variable ISSUE = 0, VALID = $k$;

function TAKE_NEXT_TICKET: ticket;
begin
    ISSUE := ISSUE + 1;
    return ISSUE;
end;

function IS_VALID($T$): Boolean
begin
    return ($T \leq$ VALID)
end;

procedure VALIDATE_NEXT_TICKET($T$);
begin
    VALID := VALID + 1;
end;
```

Fig. 3. Numbered ticket algorithm.

The code for the Numbered Ticket Algorithm is shown in Figures 2 and 3. The initial value of the local variable TICKET (Figure 2) does not matter to the operation of the algorithm.

The drawback to the Numbered Ticket Algorithm is, of course, that ISSUE and VALID grow without bound.

2.4 Colored Ticket Algorithms

We now give two variations of the Numbered Ticket Algorithm based on the idea of colored tickets, the second of which is the final Colored Ticket Algorithm.

In the previous algorithm, either ISSUE or VALID could be larger than the other, and we say the larger one *leads* the smaller. (In case of equality, each *leads* the other.) However, they could never be too far apart. If VALID leads ISSUE, then there are VALID-ISSUE valid but not-issued tickets; hence,

$$\text{VALID} - \text{ISSUE} \leq k.$$

If ISSUE leads VALID, then all k valid tickets are held by processes, and there are ISSUE − VALID invalid tickets held by processes waiting in their trying protocols. Since there are only N processes in all, k of which hold valid tickets, we conclude that

$$\text{ISSUE} - \text{VALID} \leq N - k.$$

Let $M \geq 1 + \max(k, N - k)$. Then we can determine which variable leads the other given only the information:

- $B = (\lfloor \text{VALID}/M \rfloor = \lfloor \text{ISSUE}/M \rfloor)$
- $V = \text{VALID} \bmod M$
- $I = \text{ISSUE} \bmod M$.

Namely, if B = **true**, then VALID leads ISSUE iff $V \geq I$, and if B = **false**, then VALID leads ISSUE iff $V < I$.[2] Thus, we divide the tickets into blocks of size M. B is **true** iff VALID and ISSUE are in the same block; V and I are the relative positions of VALID and ISSUE within their respective blocks. It is easy to see that if VALID and ISSUE are not in the same block, then they must be in consecutive blocks, and the condition on M insures that which block leads which can be determined by comparing V and I.

The colored ticket algorithms replace numbered tickets by *colored tickets* that consist of ordered pairs $T = (t, c)$, where t, the *value* of T, is a number between 0 and $M - 1$ indicating the position of the ticket within the block, and c, the *color* of T, is a non-negative integer indicating the block that contains the ticket. We write T.VALUE and T.COLOR to denote the two components of T. There is a natural one-to-one correspondence ψ between numbered ticket i and colored ticket $(i \bmod M, \lfloor i/M \rfloor)$. Using this correspondence, a process can determine for colored tickets whether VALID leads ISSUE without using the ordering on colors by computing:

- B := (VALID.COLOR = ISSUE.COLOR)
- V := VALID.VALUE
- I := ISSUE.VALUE

and then applying the above remarks. It also follows from the above remarks that | VALID.COLOR − ISSUE.COLOR | ≤ 1.

Now, a process can easily determine whether or not a ticket T that it holds is valid. T is always valid if its color differs from both VALID.COLOR and ISSUE.COLOR, for then its color must be less than both. If T's color is the same as VALID.COLOR, then T is valid iff T.VALUE ≤ VALID.VALUE. Finally, if T's color is the same as ISSUE.COLOR but different from VALID.COLOR, then T is valid iff VALID leads ISSUE. Using these ideas, the function IS_VALID can be defined as in Figure 4.

Unbounded Colored Ticket Algorithm. We complete the Unbounded Colored Ticket Algorithm by exhibiting in Figure 5 the definitions for the ticket issuing and validating functions. Initially, ISSUE = (0, 0) and VALID = $(k, 0)$.

The Unbounded Colored Ticket Algorithm simulates the Numbered Ticket Algorithm using the correspondence ψ between numbered tickets and colored tickets given above. Thus, we have bounded the set of ticket "values" at the cost of introducing an unbounded set of "colors." It may appear that no progress has been made, but the algorithm paves the way for the final modification which yields the space-efficient Colored Ticket Algorithm.

Colored Ticket Algorithm. We now present the main contribution of the paper, the Colored Ticket Algorithm. Like the previous algorithms, it simulates the Queue Algorithm, but it is very space efficient, requiring only $O(N^2)$ values of shared memory. It is obtained by modifying the Unbounded Colored Ticket Algorithm so that only $k + 1$ different colors are used. This requires that tickets (and colors) be reused.

[2] In case $k \neq N - k$, we can actually take $M = \max(k, N - k)$ and adjust the conditions appropriately.

```
function LEADS($A, B$): Boolean; { *Tests if $A$ leads $B$* }
begin
    if $A$.COLOR = $B$.COLOR then
        return ($A$.VALUE $\geq$ $B$.VALUE)
    else
        return ($A$.VALUE $<$ $B$.VALUE);
end;

function IS_VALID($T$): Boolean;
begin
    if $T$.COLOR = VALID.COLOR then
        return ($T$.VALUE $\leq$ VALID.VALUE)
    else if $T$.COLOR = ISSUE.COLOR then
        return LEADS(VALID, ISSUE)
    else
        return true;
end;
```

Fig. 4. Validity testing functions for colored tickets.

The change from the unbounded version of the algorithm comes when ISSUE.COLOR or VALID.COLOR is to be incremented. The new algorithm instead considers two cases. If the leading pointer (ISSUE or VALID) is being incremented, then a new color is chosen that is different from the color of any currently issued or validated ticket and different from the color of the other pointer. This insures that no two processes ever simultaneously hold the same ticket. If the trailing pointer is being incremented, then it is set equal to the color of the leading pointer. That this is correct follows from the fact that the pointers (in the Numbered Ticket Algorithm) never differ by more than M.

To see that it is always possible to select a new color when needed, we show (for the Unbounded algorithm) that every color in use *at the time a new color is needed* is the same as the color of some valid ticket; hence, at most k colors are then in use. A color is *in use* if it is the color of a valid or issued ticket that has not been discarded, or if it is equal to VALID.COLOR or ISSUE.COLOR. Note that in the exit protocol, a new ticket is validated immediately before the old one is discarded, so except for the brief moment between validating the new ticket and discarding the old one, exactly k tickets are valid, the most recently validated ticket T is still valid, and VALID.COLOR = T.COLOR. Hence, VALID.COLOR is always the color of one of the k valid tickets, so it suffices to show that when a new color is needed, both ISSUE.COLOR and the colors of all issued but not yet validated tickets are the same as VALID.COLOR.

There are two cases. If a new color is needed because VALID is about to be incremented, then VALID.VALUE = $M - 1$, VALID leads ISSUE, and a process is in its exit protocol attempting to validate a new ticket. Then ISSUE.COLOR = VALID.COLOR since ψ^{-1}(VALID) $-$ ψ^{-1}(ISSUE) $\leq k \leq M - 1$. Since there

```
constant $M = 1 + \max(k, N - k)$;

global variable ISSUE = (0,0), VALID = (k,0)

function TAKE_NEXT_TICKET: ticket;
begin
    if ISSUE.VALUE < $M - 1$ then
        ISSUE.VALUE := ISSUE.VALUE + 1
    else begin
        ISSUE.VALUE := 0;
        ISSUE.COLOR := ISSUE.COLOR + 1;
        end;
    return ISSUE;
end;

procedure VALIDATE_NEXT_TICKET(T);
begin
    if VALID.VALUE < $M - 1$ then
        VALID.VALUE := VALID.VALUE + 1
    else begin
        VALID.VALUE := 0;
        VALID.COLOR := VALID.COLOR + 1;
        end;
end;
```

Fig. 5. Unbounded colored ticket algorithm.

are no issued but not validated tickets, the only colors in use are those belonging to the k valid tickets.

On the other hand, if a new color is needed because ISSUE is about to be incremented, then ISSUE.VALUE = $M - 1$, ISSUE leads VALID, and a ticket is about to be issued to an entering process. Again, ISSUE.COLOR = VALID.COLOR, for $\psi^{-1}(\text{ISSUE}) - \psi^{-1}(\text{VALID}) \le N - k \le M - 1$.[3] Moreover, any outstanding invalid tickets lie between VALID and ISSUE, so they also have color VALID.COLOR. Again the only colors in use are those belonging to the k valid tickets.

We conclude that with $k + 1$ colors altogether, there is always a free color whenever a new one is needed.

To permit a process to determine which color is free, we introduce an array QUANT of length $k + 1$ into the shared variable, where QUANT(c) $\in \{0, 1, \ldots, k\}$ gives the number of valid tickets of color c. There are exactly k valid tickets, so the total number of different values for the QUANT array is the number of $(k + 1)$-tuples of non-negative integers that sum to k. This in turn is the number

[3] Actually, $\psi^{-1}(\text{ISSUE}) - \psi^{-1}(\text{VALID}) \le N - k - 1$ since the entering process does not yet hold a ticket, but we do not make use of this fact.

```
function NEW_COLOR: integer; { Returns unused color }
local variable $C$;
begin
    $C := 0$;
    while QUANT($C$) > 0 do $C := C + 1$;
    return $C$
end;
```

Fig. 6. Find unused color function (used by colored ticket algorithm).

of partitions of k identical elements into $k + 1$ sets, or $\binom{2k}{k}$. While this number is exponential in k, it is independent of N. QUANT is updated whenever a ticket is discarded and a new one is validated.

The code for finding a new color is shown in Figure 6. It simply scans for a color with QUANT = 0. Although NEW_COLOR appears to involve a loop which takes time $O(k)$, only constant time is required in our model of computation. We could make this explicit at the cost of a factor k^k additional values in shared memory. The idea is to keep a stack of indices of QUANT elements that have value zero. NEW_COLOR then simply pops an index from the stack. VALIDATE_NEXT_TICKET checks to see if QUANT(T.COLOR) is zero and, if so, pushes it onto the stack.

The final algorithm is contained in Figures 2, 4, 6, and 7. Initially, ISSUE = (0, 0) and VALID = (k, 0).

This algorithm simulates the Unbounded Colored Ticket Algorithm. To prove this, one shows that any two issued or validated (and not discarded) tickets T and T' have the same color in this algorithm iff they have the same color in the Unbounded algorithm; hence, the two algorithms always make the same decisions. We leave the details to the reader.

The total number of shared memory values needed by the Colored Ticket Algorithm is the product of the number of values assumed by QUANT, ISSUE, and VALID. This works out to

$$\binom{2k}{k}((k+1)M)^2 = O(N^2)$$

as desired, since $M = O(N)$.

3. A FORMAL MODEL FOR SYSTEMS OF PROCESSES

We now present a formal model of computation and state the conditions that define the k-critical section problem. The model is derived from that of [2]. It can also be regarded as a special case of the general model of [16].

3.1 Processes and Systems

A *process* is a quadruple $P = (V, X, \delta, \mathscr{R})$, where

- V is a set of values for a shared variable,
- X is a (not necessarily finite) set of process states,

```
constant $M = 1 + \max(k, N - k)$;

global variable ISSUE = (0,0), VALID = ($k$,0), QUANT[0],...,QUANT[$k$] = 0;

function TAKE_NEXT_TICKET: ticket;
begin
    if ISSUE.VALUE < $M - 1$ then
        ISSUE.VALUE := ISSUE.VALUE + 1
    else begin
        if LEADS(ISSUE, VALID) then
            ISSUE.COLOR := NEW_COLOR
        else
            ISSUE.COLOR := VALID.COLOR;
        ISSUE.VALUE := 0
        end;
    return ISSUE;
end;

procedure VALIDATE_NEXT_TICKET($T$);
begin
    if VALID.VALUE < $M - 1$ then
        VALID.VALUE := VALID.VALUE + 1
    else begin
        if LEADS(VALID, ISSUE) then
            VALID.COLOR := NEW_COLOR
        else
            VALID.COLOR := ISSUE.COLOR;
        VALID.VALUE := 0
        end;

    { Update quantity information. }

    QUANT(VALID.COLOR) := QUANT(VALID.COLOR) + 1;
    QUANT($T$.COLOR) := QUANT($T$.COLOR) - 1;
end;
```

Fig. 7. Colored ticket algorithm.

- δ is a total function from $V \times X$ to $V \times X$, the *transition function*, and
- $\mathscr{R}$ is a total function from X to $\{R, T, C, E\}$, the *region function*.

Assume process P is in state x and the shared memory has value v. A *step* of P changes the state to x' and the shared memory to v', where $(v', x') = \delta(v, x)$.

For a state $x \in X$, $\mathscr{R}(x)$ gives the *region* of x, where R denotes the *remainder region*, T the *trying region*, C the *critical region*, and E the *exit region*. We assume

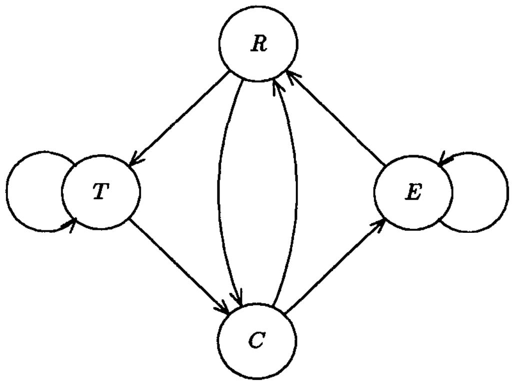

Fig. 8. Possible region changes.

that δ respects $\mathscr{R}$ as follows. For every $(v, x) \in V \times X$:

(1) $\mathscr{R}(x) \in \{R, T\}$ implies $\mathscr{R}(\delta(v, x)) \in \{T, C\}$, and
(2) $\mathscr{R}(x) \in \{C, E\}$ implies $\mathscr{R}(\delta(v, x)) \in \{E, R\}$.

The allowed transitions are indicated in Figure 8. The transitions out of R and T comprise the trying protocol, and the transitions out of C and E comprise the exit protocol.[4] We "abstract away" the steps comprising the critical and remainder sections treating only the protocols explicitly; hence the absence of self-loops on R and C. Thus the next step of a process in R takes it out of R, and similarly for C.

For a natural number N, let $[N]$ denote $\{1, \ldots, N\}$. A *system S of N processes* is a collection of processes $P_i = (V, X_i, \delta_i, \mathscr{R}_i)$, $i \in [N]$, all having the same shared variable V.

An *instantaneous description* (i.d.) q of S is a snapshot of the configuration of S and completely determines S's possible future behaviors. Formally q is an $(N + 1)$-tuple $(v, x_1, \ldots, x_N)$, where $v \in V$ is the contents of the shared variable and $x_i \in X_i$, $1 \le i \le N$, are the states of the N processes. We denote v by $V(q)$ and x_i by $X_i(q)$, $1 \le i \le N$.

The functions δ_i and $\mathscr{R}_i$ of the individual processes are naturally extended to functions on the set of i.d.'s of S by defining

$$\delta_i(v, x_1, \ldots, x_N) = (v', x_1, \ldots, x_{i-1}, x', x_{i+1}, \ldots, x_N)$$

where $(v', x') = \delta_i(v, x_i)$, and

$$\mathscr{R}_i(v, x_1, \ldots, x_N) = \mathscr{R}_i(x_i).$$

[4] Our formal model imposes a slight restriction on the form of protocols in that all transitions leaving a state of the trying region must belong to the trying protocol (and similarly for the exit region and protocol). Thus a process, once permitted to begin its critical section, must first take a step to leave the trying region before it begins executing steps of its critical section, and the step which takes it out of the trying region is considered to be a part of the trying protocol. This restriction is for technical convenience only and does not weaken the results, for any protocol can be easily put into this form by adding dummy steps to the ends of the trying and exit protocols.

A *schedule* h for S is any finite or infinite sequence of elements of $[N]$.[5] A schedule describes the interleaving of process steps in a particular "run" of the system. Since the processes are deterministic, the entire run is determined by the starting i.d. q of the system and a schedule h. Formally the *run determined by q and $h = h_1, h_2, \ldots$* is the finite or infinite sequence of i.d.'s $Q(q, h) = q_0, q_1, q_2, \ldots$ such that:

(1) If h is infinite then $Q(q, h)$ is infinite, and if h is finite then $|Q(q, h)| = |h| + 1$.

(2) $q_0 = q$.

(3) If q_{i-1}, q_i are successive elements of $Q(q, h)$, then $q_i = \delta_{h_i}(q_{i-1})$.

If $Q(q, h)$ is finite, then the last i.d., q_s, is the *result* of $Q(q, h)$, and we denote q_s by $\delta(q, h)$, extending δ once again. I.d. q' is *reachable from q via h* provided $\delta(q, h) = q'$, and q' is *reachable from q* if q' is reachable from q via some finite schedule h.

3.2 Equivalence of Systems

Let S and S' be systems of N processes, with q and q' i.d.'s of S and S' respectively. We say that (S, q) and (S', q') are *equivalent* if for every finite schedule h, all processes are in the same regions in $\delta(q, h)$ and $\delta'(q', h)$; that is, for every $i \in [N]$, $\mathscr{R}_i(\delta(q, h)) = \mathscr{R}'_i(\delta'(q', h))$.

3.3 Dependencies Among Processes

We have noted that processes are always free to leave their remainder or critical regions on their own, but the same is not true for the trying and exit regions. We next give some important definitions that describe possible dependencies among processes progressing through their regions.

Let Z denote any region. A process P_i in a system of processes is *Z-enabled* in i.d. q if for every schedule $\bar{h}$ in which i occurs infinitely often, there is a finite prefix h of $\bar{h}$ such that $\mathscr{R}_i(\delta(q, h)) = Z$. Thus the Z-enabled i.d.'s are those in which a process is either already in Z or will eventually enter Z (if it tries long enough), no matter what the other processes do. Note that a process P_i can become Z-enabled because of its own actions or because of actions of other processes. Thus P_i might not be Z-enabled in q and yet be Z-enabled in $\delta(q, h)$, even if h does not contain i. In this case we say that P_i has been *passively enabled from q by h* and can be thought of as passively belonging to region Z.

We say that P_i is *T-waiting* in q if it is in T but is not C-enabled in q. Similarly, we say that P_i is *E-waiting* in q if it is in E but is not R-enabled in q.

3.4 Properties of Systems

We now state the properties that define the k-critical section problem. Throughout this section, S denotes a system of N processes, q an i.d. of S, $k < N$ a natural number, and $\#Y$ the cardinality of the set Y.

[5] Note that h is *not* required to be "fair." Processes that take only finitely many steps in h are considered to have failed.

Our first condition is the basic k-exclusion condition.

- **k-Exclusion.** I.d. q satisfies *k-exclusion* if $\#\{i \in [N] \mid \mathscr{R}_i(q) = C\} \le k$. S satisfies *k-exclusion from* q if every i.d. reachable from q in S satisfies k-exclusion.

Note that any set of processes that are C-enabled but not in C can, by taking steps on their own, reach an i.d. in which all are simultaneously in C. Thus, if S satisfies k-exclusion from q, the number of C-enabled processes in any i.d. reachable from q is at most k.

Our second condition describes our robustness requirements. We say that i.d. q is *k-full* if $\#\{i \in [N] \mid P_i$ is C-enabled in $q\} \ge k$. We say that a process P_i *makes progress* in a run if, for some pair of i.d.'s q' and q'' in the run, either

(1) $\mathscr{R}_i(q') \ne \mathscr{R}_i(q'')$, or
(2) P_i is T-waiting in q' but not in q'', or
(3) P_i is E-waiting in q' but not in q''.

- **Avoidance of k-Deadlock.** An infinite schedule h *exhibits k-deadlock from* q if no process makes progress in the run $Q(q, h)$, and either

  (1) some process is T-waiting in q and q is not k-full, or
  (2) some process is E-waiting in q.[6]

  S avoids k-deadlock from q if no infinite schedule exhibits k-deadlock from any i.d. reachable from q.

Our third and final condition describes the fairness property, FIFO enabling. Intuitively, violation of FIFO enabling occurs if a process remains T-waiting while another process, beginning in its remainder region, becomes C-enabled. Similarly, a violation occurs if a process remains E-waiting while another process, beginning in its critical region, becomes R-enabled. Formally let q be an i.d. and h a finite schedule. We say P_j *overtakes* P_i in $Q(q, h)$ if P_i is T-waiting in all i.d.'s of $Q(q, h)$, $\mathscr{R}_j(q) = R$, and P_j is C-enabled in $\delta(q, h)$, or if P_i is E-waiting in all i.d.'s of $Q(q, h)$, $\mathscr{R}_j(q) = C$, and P_j is R-enabled in $\delta(q, h)$.

- **FIFO Enabling.** *S achieves FIFO enabling* from q if for all q' reachable from q, all finite schedules h, and all $i, j \in [N]$, P_j does not overtake P_i in $Q(q', h)$.

The Problem. Let q be an i.d. with every process in its remainder region. A system S solves the *k-critical section problem* starting from q if it satisfies k-exclusion, avoids k-deadlock, and achieves FIFO enabling from q.

[6] Intuitively a schedule exhibits k-deadlock if some process "wants" to make progress and progress is possible, but no process actually does make progress. At first sight, it might seem necessary to exclude failed processes from consideration in the formal definition, for we do not consider that progress is possible for failed processes. However it is unnecessary to distinguish between failed and nonfailed processes because our convention of no self-loops on R and C implies that *every* nonfailed process "wants" to make progress (since it cannot continue taking steps and remain in R or C), and at least one process is nonfailed in every infinite schedule.

4. UPPER BOUND

The Colored Ticket Algorithm, when translated into the formalism of our model, shows that the k-critical section problem can be solved by a system S that uses only $O(N^2)$ values of shared memory.

The translation requires a few comments. The atomic actions used in the algorithm make several accesses to the shared global variables, change internal variables, and branch to one of several possible exits depending on the values in shared and private memory at the start of the action. In our formal model, each atomic action becomes a single process step. The program counter and all internal storage of a process is represented by the state x, and the entire contents of the global variables is represented by the value v of the shared variable. To construct the value (v', x') of the transition function $\delta(v, x)$, if the program counter in x points to a **start** instruction, then run the algorithm until it encounters a **finish** statement, and move the program counter past the **finish**. x' is the state and v' the shared memory contents that result. If a **finish** is never reached, or if the program counter in x does not point to a **start** instruction, then $\delta(v, x)$ is defined arbitrarily. This translation is not fully general, but it is adequate for algorithms such as ours in which every atomic action terminates, and the next instruction to be executed after a **finish** is always a **start**.

THEOREM 4.1. *The Colored Ticket Algorithm, when translated into the formal model as described above, solves the k-critical section problem and uses*

$$\binom{2k}{k}((k+1)(1+\max(k, N-k)))^2 = O(N^2)$$

values of shared memory.

A formal proof can be constructed following the development given in Section 2. Namely, one first proves that the Queue Algorithm solves the k-critical section problem. Next one shows that each of the three successively presented algorithms is equivalent to the preceding in the sense formally defined in Section 3.2. Finally one applies the following lemma, whose proof is straightforward.

LEMMA 4.2. *Assume (S, q) is equivalent to (S', q'). If S satisfies the k-critical section problem from q, then S' satisfies the k-critical section problem from q'.*

5. LOWER BOUND

In this section we establish a lower bound on the size of the shared variable of any system of processes that solves the k-critical section problem. We assume throughout that k and N are natural numbers with $N \geq k + 2$, S is a system of N processes, and q_0 is an i.d. with every process in its remainder region such that S solves the k-critical section problem from q_0.

Our method of proof is to construct a collection of runs and show that each leaves the shared variable in a distinct state. In order to carry out the construction, we need several "liveness" lemmas that show certain kinds of progress are always possible.

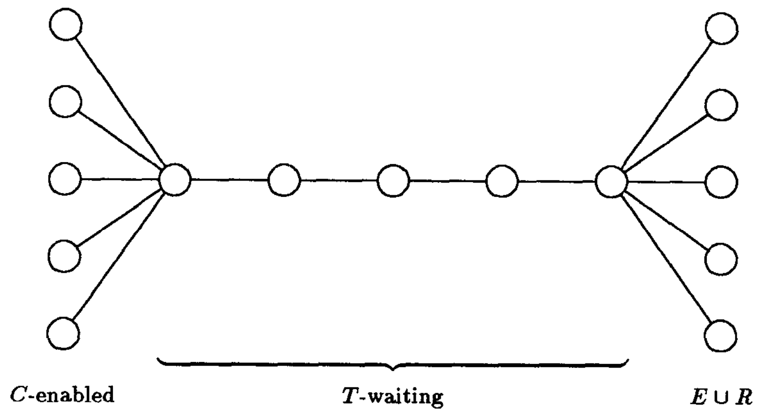

Fig. 9. The relation $\prec_q$.

5.1 Progress Lemmas

We begin with some basic properties that follow from the fact that S solves the k-critical section problem. FIFO enabling places rather severe constraints on the order in which processes can become C-enabled, which are expressed by the relation $\prec_q$ that we next define.

Consider any i.d. q and processes P_i and P_j. We define $i \prec_q j$ to hold precisely if one of the following conditions holds at q:

(1) P_i is C-enabled and P_j is in $E \cup R$;
(2) P_i is C-enabled and P_j is T-waiting;
(3) P_i is T-waiting and P_j is in $E \cup R$;
(4) P_i and P_j are both T-waiting, and in some run leading from q_0 to q, P_i last entered T before P_j did.

We also define $\text{ahead}_j(q) = \{i \in [N] \mid i \prec_q j\}$. The ordering $\prec_q$ is illustrated in Figure 9.[7]

The first lemma says that the order in which processes become C-enabled from q respects $\prec_q$.

LEMMA 5.1. *Let q be reachable from q_0, and let $i \prec_q j$. Let h be a finite schedule such that P_j is C-enabled in $\delta(q, h)$. Then P_i is C-enabled in some i.d. in $Q(q, h)$.*

PROOF. Assume the conditions of the lemma. Since $i \prec_q j$, P_i is either C-enabled or T-waiting in q. If P_i is C-enabled, then we simply choose $h' = \lambda$, the null schedule, and we are done. Hence, assume P_i is T-waiting in q.

Again, since $i \prec_q j$, P_j is either in $E \cup R$ or is T-waiting in q. In either case, there exists an i.d. q_1 (possibly equal to q) and schedules h_0, h_1 such that q_1 is reachable from q_0 via h_0, q is reachable from q_1 via h_1, P_j is in $E \cup R$ in q_1, and P_i is T-waiting in every $q' \in Q(q_1, h_1)$. P_i is not T-waiting in every $q' \in Q(q, h)$,

[7] One can show that if q is reachable from q_0, then $\prec_q$ is a strict partial order which totally orders the T-waiting processes in q, as illustrated, but we do not need this fact.

for if it were, then P_j overtakes P_i in $Q(q_1, h_1 \cdot h)$, violating FIFO enabling. Hence, P_i is C-enabled in some i.d. in $Q(q, h)$. □

The next lemma implies that among the T-waiting processes there is one that is "ahead" of all the others.

LEMMA 5.2. *Let q be reachable from q_0, and assume that at least one process is T-waiting in q. Then there is a T-waiting process P_i in q such that $i \prec_q j$ for all $j \neq i$ such that P_j is T-waiting in q.*

PROOF. Let q be reachable from q_0 via h, and consider the run $Q(q_0, h)$. Order the T-waiting processes in q according to the times of their most recent entry to T in $Q(q_0, h)$, and let P_i be the first such process. By definition, $i \prec_q j$ holds for all $j \neq i$ such that P_j is T-waiting in q. □

LEMMA 5.3. *Let q be reachable from q_0.*

(1) *If q is not k-full, then no process is T-waiting in q.*
(2) *No process is E-waiting in q.*

PROOF. (1) Assume that q is not k-full but some process is T-waiting in q. We proceed to derive a contradiction.

By Lemma 5.2, there is a T-waiting process P_i in q such that $i \prec_q j$ for all $j \neq i$ such that P_j is T-waiting in q. Since P_i is T-waiting in q, there is a schedule h in which P_i takes infinitely many steps but it remains in T in every i.d. of $Q(q, h)$; hence P_i is T-waiting in every i.d. of $Q(q, h)$.

Suppose a process P_j becomes C-enabled during $Q(q, h)$. That is, suppose one can write $h = h_1 \cdot h_2 \cdot h_3$ such that P_j is not C-enabled in $q_1 = \delta(q, h_1)$, but P_j is C-enabled in $\delta(q_1, h_2)$. Then $i \prec_{q_1} j$ holds by definition, so by Lemma 5.1, P_i is C-enabled at some i.d. in $Q(q_1, h_2)$, a contradiction. Hence, no process becomes C-enabled during $Q(q, h)$, so none of the i.d.'s in $Q(q, h)$ are k-full.

Now, for some suffix $Q(q', h')$ of the run $Q(q, h)$, no process makes progress since each process can change region or become R-enabled only a finite number of times without becoming C-enabled. Thus, h' exhibits k-deadlock from q', contradicting the avoidance of k-deadlock condition.

(2) The proof is similar (and simpler). Assume that P_i is E-waiting in q. Then there is a schedule h in which P_i takes infinitely many steps but it remains in E in every i.d. of $Q(q, h)$. It follows that P_i is E-waiting in every i.d. of $Q(q, h)$.

Only processes P_j already in E in q can become R-enabled during $Q(q, h)$, and that can happen at most once per process, for otherwise P_j would overtake P_i, violating the FIFO enabling condition. Hence, in some suffix $Q(q', h')$ of the run $Q(q, h)$ no process makes progress since each process can change region or become C-enabled only a finite number of times without becoming R-enabled. Then h' exhibits k-deadlock from q', contradicting the avoidance of k-deadlock condition. □

The following lemma says that a process can only be C-enabled while it is in its trying or critical region.

LEMMA 5.4. *Let q be reachable from q_0. If process P_i is C-enabled in q, then P_i is in $T \cup C$ in q.*

PROOF. Assume the contrary, that P_i is C-enabled in q, and P_i is in $E \cup R$ in q. For each $j \in [N]$, $j \neq i$, run P_j for zero or more steps until an i.d. is reached in which it is in $T \cup C$. This procedure must terminate after a finite number of steps, for otherwise P_j remains forever in $E \cup R$. But that is impossible by Lemma 5.3 and the absence of self-loops on region R. Call the resulting i.d. q'.

In q', every process other than P_i is either T-waiting or is C-enabled. At most k processes can be C-enabled (by the remark following the definition of k-exclusion). Thus, since we assume $N \geq k + 2$, some process P_ℓ is T-waiting in q'. P_i is still C-enabled in q' (by definition of enabling), so it enters C in the run $Q(q', i^m)$ for some m. By Lemma 5.3, q' is k-full, so P_ℓ remains T-waiting throughout $Q(q', i^m)$. But then P_i overtakes P_ℓ in $Q(q', i^m)$, violating FIFO enabling. □

The next lemma says that, no matter what the other processes do, any process in its trying region that takes infinitely many steps eventually reaches its critical region, provided that there are not too many processes ahead of it.

LEMMA 5.5. *Let q be reachable from q_0, and let P_i be in T in q. Then $\#ahead_i(q) < k$ iff P_i is C-enabled in q.*

PROOF. Assume the conditions of the lemma.

($\Rightarrow$) Suppose $\#\text{ahead}_i(q) < k$ but P_i is not C-enabled in q. Then P_i must be T-waiting in q, so by Lemma 5.3, q is k-full. But then all the processes that are C-enabled in q are in $\text{ahead}_i(q)$, so that $\#\text{ahead}_i(q) \geq k$. This is a contradiction.

($\Leftarrow$) If P_i is C-enabled in q, then P_i is in $T \cup C$ by Lemma 5.4. But then $\text{ahead}_i(q) = \varnothing$. □

The next lemma says that it is always possible for all the processes to run so as to end up simultaneously in their remainder regions.

LEMMA 5.6. *Let q be reachable from q_0. Then there exists q' reachable from q such that every process is in its remainder region in q'. Moreover, q' can be reached from q via a schedule in which no process already in its remainder region in q takes any steps.*

PROOF. It suffices to show that if not all processes are in their remainder regions, then there is some P_i not in R that is C- or R-enabled. Assuming we have shown that such a P_i exists, we run P_i until it changes regions. We then repeat this construction on each resulting i.d. until an i.d. is reached in which all processes are in R. This procedure must eventually terminate since each process can change regions only finitely many times before entering its remainder region.

Now suppose that every process not in R is neither C- nor R-enabled in q. Then q is not k-full, since no process is C-enabled, by assumption and Lemma 5.4. By Lemma 5.3, no process is T- or E-waiting in q; therefore, no process is in $T \cup E$ in q. But also no process is in C in q since no process is C-enabled. Hence every process is in R. It follows that if not all processes are in R, then some such process is C- or R-enabled, as desired. □

$$\begin{tikzcd}
P_1 & \underbrace{P_2 \ldots P_k} & \underbrace{P_{k+1} \ldots P_j} & \underbrace{P_{j+1} \ldots P_i} & \underbrace{P_{i+1} \ldots P_N} & I.D. \\
R \arrow[d, "*"] & R \arrow[d, "*"] & R \arrow[d, "1"] & R \arrow[d, "1"] & R \arrow[d, "1"] & q \\
C \arrow[d, "*"] & C \arrow[d, "0"] & T \arrow[d, "0"] & T \arrow[d, "0"] & T \arrow[d, "0"] & \\
R \arrow[d, "0"] & C \arrow[d, "0"] & T \arrow[d, "*"] & T \arrow[d, "*"] & T \arrow[d, "0"] & q' \\
R \arrow[d, "0"] & C \arrow[d, "0"] & R \arrow[d, "1"] & R \arrow[d, "0"] & T \arrow[d, "0"] & \\
R & C & T & R & T & q(i,j)
\end{tikzcd}$$

Fig. 10. The lower bound construction.

5.2 The Schedule $h(i, j)$

Now choose any q reachable from q_0 in which all processes are in their remainder regions. q exists by Lemma 5.6. Fix i and j, with $k \leq j < i \leq N - 1$. Construct a schedule, $h(i, j)$, as follows.

(1) Starting at q, each of $P_1, \ldots, P_k$ takes steps on its own, just until it enters its critical region. This is possible by Lemma 5.5. Then each of $P_{k+1}, \ldots, P_N$ takes one step, going to its trying region. Let P_N's state after its entry be denoted by x—for future reference. (Note that x does not depend on i or j.)

(2) P_1 takes steps on its own, just until it returns to its remainder region, leaving one empty critical slot. This is possible by Lemma 5.3. Call the resulting i.d. q' for later reference. (Note that q' does not depend on i or j.)

(3) Each of $P_{k+1}, \ldots, P_i$ in turn takes steps on its own, just until it returns to its remainder region. This is possible by Lemmas 5.3 and 5.5.

(4) Each of $P_{k+1}, \ldots, P_j$ takes one step, thereby entering its trying region once again. The resulting i.d. is denoted $q(i, j)$.

This construction is diagrammed in Figure 10. Arrows are labeled by '0', '1', or '*' to indicate that the corresponding process takes 0, 1, or an unspecified number of steps.

5.3 Distinctness of Shared Values

We now relate the construction to the size of shared memory.

LEMMA 5.7. *The shared variable has a distinct value in each* $q(i, j)$.

PROOF. Assume to the contrary that $V(q(i, j)) = V(q(i', j'))$ for $(i, j) \neq (i', j')$. Without loss of generality, it suffices to consider two cases.

Case 1. $i < i'$. Among all T-waiting processes in $q(i',j')$, $P_{i'+1}$ was the first to enter its trying region, so $\#\text{ahead}_{i'+1}(q(i',j')) = k - 1$. Then $P_{i'+1}$ is C-enabled in $q(i',j')$ by Lemma 5.5. Also, $P_{i'+1}$ is in the same state in both $q(i,j)$ and $q(i',j')$. Since also the shared variable has the same value in both i.d.'s, it follows that $P_{i'+1}$, starting from $q(i,j)$ can take some number m of steps and enter its critical region. We claim that the schedule $\bar{h} = h(i,j) \cdot (i' + 1)^m$ violates FIFO enabling from q. This is because $P_{i'+1}$ goes from its remainder to its critical region during $\bar{h}$ while P_{i+1}, which entered its trying region first, remains T-waiting. (P_{i+1} does not become C-enabled during $\bar{h}$, for if it did, then k-exclusion would be violated in the schedule $\bar{h} \cdot (i + 1)^\omega$.)

Case 2. $i = i'$ and $j < j'$. Consider schedule h constructed as follows. Starting from $q(i,j)$, $P_{j'+1}$ takes one step, thereby entering the trying region. Then each of $P_{i+1}, \ldots, P_N, P_{k+1}, \ldots, P_j$, in turn, takes sufficiently many steps to return to its remainder region, possible by Lemmas 5.3 and 5.5. Call the resulting i.d. q_1. Then $\text{ahead}_{j'+1}(q_1) = \{2, \ldots, k\}$, so $P_{j'+1}$ is C-enabled in q_1 by Lemma 5.5.

Now consider the application of h to $q(i,j')$ and let q_1' be the resulting i.d. $P_{j'+1}$ is in the same state in both q_1 and q_1', and also the shared variable has the same value in both i.d.'s; thus $P_{j'+1}$, starting from q_1', can take some number m of steps and enter its critical region. Hence, $P_{j'+1}$ enters its critical region in the run $Q(q(i,j'), h')$, where $h' = h \cdot (j' + 1)^m$. The schedule h' violates FIFO enabling from $q(i,j')$, for $P_{j'+1}$ overtakes P_{j+1} in $Q(q(i,j'), h')$. □

5.4 Lower Bound Theorem

Finally we prove the main lower bound result.

THEOREM 5.8. *Let $N \geq k + 2$, and let S be a system of N processes with value set V for its shared variable, and let q_0 be an i.d. such that S solves the k-critical section problem from q_0. Then,*

$$|V| \geq k\binom{N-k-1}{2} + N - k - 1 = \Omega(N^2).$$

PROOF. The proof proceeds by induction on k.

Base. $k = 1$. By Lemma 5.7, there are at least $\binom{N-1}{2} = 1 \cdot \binom{N-2}{2} + N - 2$ distinct values.

Inductive step. $k > 1$. By Lemma 5.7, there are $\binom{N-k-1}{2}$ distinct values of the variable for the i.d.'s $q(i,j)$ for i, j satisfying $k \leq j < i \leq N - 2$. Each such $q(i,j)$ is k-full since $P_2, \ldots, P_k$ and P_{i+1} are C-enabled in $q(i,j)$. Hence, by k-exclusion, if $v(i,j)$ is the value of the shared variable in $q(i,j)$, then no finite number of applications of P_N's transition function δ_N to the pair $(v(i,j), x)$ can put P_N in its critical region. (Recall that P_N is in the same state x in each $q(i,j)$.)

Now reconsider the construction of Section 5.2. Starting at q', let each of $P_2, \ldots, P_k$ take steps until they return to their remainder regions, possible since all are R-enabled by Lemma 5.5. Now let each of $P_{k+1}, \ldots, P_{N-1}$ in turn enter their critical regions and then return to their remainder regions, again possible by Lemmas 5.3 and 5.5. Call the resulting i.d. q''.

$P_1, \ldots, P_{N-1}$ are in their remainder regions and P_N is in its trying region in q'', so P_N is C-enabled in q'' by Lemma 5.5. From q'', consider $P_1, \ldots, P_{N-1}$ as comprising a system, S', of $N - 1$ processes. Since S solves the k-critical section problem from q, it can be shown that S' solves the $(k - 1)$-critical section problem from (the appropriate restriction of) q''. Thus, by induction, the number of values that can be taken on by S''s shared variable is at least

$$\begin{aligned}(k-1)&\binom{(N-1)-(k-1)-1}{2} + (N-1)-(k-1)-1\\ &= (k-1)\binom{N-k-1}{2} + N-k-1.\end{aligned}$$

Since P_N is C-enabled in q'', each value v that can be taken on by the shared variable in i.d.'s reachable from q'' using only $P_1, \ldots, P_{N-1}$ has the property that some finite number of applications of δ_N to the pair (v, x) will put P_N in its critical region. Thus these shared variable values are disjoint from the values $v(i, j)$ considered above.

We conclude that

$$\begin{aligned}|V| &\geq \binom{N-k-1}{2} + (k-1)\binom{N-k-1}{2} + N-k-1\\ &= k\binom{N-k-1}{2} + N-k-1,\end{aligned}$$

as desired. □

6. SUMMARY AND OPEN QUESTIONS

In this paper, we have described the k-critical section problem in general terms and have defined an extremely robust version of the problem: equivalence with a particular simple, but space-inefficient, algorithm, the Queue Algorithm.

As our main result, we have presented an interesting new algorithm, the Colored Ticket Algorithm, which solves the given version of the problem and uses only $O(N^2)$ values of the shared variable. Our lower bound proof shows that, for fixed k, this algorithm is optimal to within a constant factor in terms of number of values of shared memory.

There is still a large gap between the constants in the upper and lower bounds. Both depend on k, but the constant in the upper bound is exponential in k, while the constant in the lower bound is linear in k. It remains to close this gap.

Our results assume that k is fixed. If we allow k to be a function of N our algorithms are not tight. For example, if $k = \log N$, then our algorithm uses $O(N^4(\log N)^{3/2})$ values of shared memory, but our lower bound is only $\Omega(N^2 \log N)$. If $k = O(N)$, then the gap is exponential. These cases deserve investigation.

ACKNOWLEDGMENTS

We are grateful to Brian Coan for helpful comments on an early draft of this paper. We are also grateful to the referees for many helpful suggestions for improving the presentation.

REFERENCES

1. BURNS, J. E. Complexity of communication among asynchronous parallel processes. Ph.D. dissertation, School of Information and Computer Science, Georgia Institute of Technology, 1981.
2. BURNS, J. E., JACKSON, P., LYNCH, N. A., FISCHER, M. J., AND PETERSON, G. L. Data requirements for implementation of N-process mutual exclusion using a single shared variable. *J. ACM 29*, 1 (1982), 183–205.
3. CREMERS, A. B., AND HIBBARD, T. N. An algebraic approach to concurrent programming control and related complexity problems. Tech. Rep., University of Southern California, Nov. 1975. (Presented at Symposium on Algorithms and Complexity, Pittsburgh, Pa., April 1976.)
4. CREMERS, A. B., AND HIBBARD, T. N. Mutual exclusion of N processors using an $O(N)$-valued message variable. In *Proceedings of 5th ICALP* (Udine, Italy). Lecture Notes in Computer Science, vol. 62. Springer Verlag, New York, pp. 165–176.
5. CREMERS, A. B., AND HIBBARD, T. N. Arbitration and queueing under limited shared storage requirements. Tech. Rep. 83, Dept. of Informatics, University of Dortmund, Mar. 1979.
6. DEBRUIJN, N. G. Additional comments on a problem in concurrent control. *Commun. ACM 10*, 3 (Mar. 1967), 137–138.
7. DIJKSTRA, E. W. Solution of a problem in concurrent programming control. *Commun. ACM 8*, 9 (1965), 569.
8. EISENBERG, M. A., AND MCGUIRE, M. R. Further comments on Dijkstra's concurrent programming control problem. *Commun. ACM 15*, 11 (Nov. 1972), 999.
9. FISCHER, M. J., LYNCH, N. A., BURNS, J. E., AND BORODIN, A. Resource allocation with immunity to limited process failure. In *Proceedings of 20th Annual IEEE Symposium on Foundations of Computer Science* (San Juan, P.R., Oct. 1979), IEEE, New York, pp. 234–254.
10. KNUTH, D. E. Additional comments on a problem in concurrent programming control. *Commun. ACM 9*, 5 (1966), 321–322.
11. LAMPORT, L. A new solution of Dijkstra's concurrent program problem. *Commun. ACM 17*, 8 (1974), 453–455.
12. LAMPORT, L. The synchronization of independent processes. *Acta Inf. 7* (1976), 15–34.
13. LAMPORT, L. A bug in the bakery algorithm. Tech. Rep. CA-7704-0611, Massachusetts Computer Associates, Inc., Apr. 1977.
14. LAMPORT, L. The mutual exclusion problem: Part I—A theory of interprocess communication. *J. ACM 33*, 2 (1986), 313–326.
15. LAMPORT, L. The mutual exclusion problem: Part II—Statement and solutions. *J. ACM 33*, 2 (1986), 327–348.
16. LYNCH, N. A., AND FISCHER, M. J. On describing the behavior and implementation of distributed systems. *Theor. Comput. Sci. 13* (1981), 17–43.
17. LYNCH, N. A., AND FISCHER, M. J. A technique for decomposing algorithms which use a single shared variable. *J. Comput. Syst. Sci. 27*, 3 (1983), 350–377.
18. MORRIS, J. M. A starvation-free solution to the mutual exclusion problem. *Inf. Process. Lett. 8*, 2 (Feb. 1979), 76-80.
19. PETERSON, G. L. New bounds on mutual exclusion problems. Tech. Rep. TR 68, University of Rochester, Rochester, N.Y., Feb. 1980.
20. PETERSON, G. L. Myths about the mutual exclusion problem. *Inf. Process. Lett. 12*, 3 (June 1981), 115–116.
21. PETERSON, G. L., AND FISCHER, M. J. Economical solutions for the critical section problem in a distributed system. In *Proceedings of Ninth ACM Symposium on Theory of Computing* (May 1977), ACM, New York, pp. 91–97.
22. RIVEST, R. L., AND PRATT, V. R. The mutual exclusion problem for unreliable processes: Preliminary report. In *Proceedings of 17th Annual IEEE Symposium on Foundations of Computer Science* (1976), IEEE, New York, 1976, 1–8.

Received July 1985; revised March 1987; accepted September 1988