

A class of allocation strategies inducing bounded delays only.

by Edsger W.Dijkstra

We consider a finite set of persons, say numbered from 1 through M, whose never ending life consists of an alternation of eating and thinking, i.e. (in the first instance) they all behave according to the program

```

cycle begin eat;
           think
end.

```

The persons are living in parallel and their common accommodations are such that not all combinations of simultaneous eaters are permitted. As a result: when a person has finished thinking, some inspection has to take place in order to decide whether he can (and shall) be granted access to the table or not. Similarly, when a person leaves the table, some inspection has to take place in order to discover whether on account of the changed occupancy of the table one or more hungry persons could (and should) be admitted to the table. This situation is reflected by writing their program

```

cycle begin ENTRY;
           eat;
           EXIT;
           think
end

```

with the understanding that

- 1) all inspection processes "ENTRY" and "EXIT" take only a finite period of time and exclude each other in time. (As a result of the postulated mutual exclusion of the inspections "ENTRY" and "EXIT", a "local delay" of person *i*, wanting to invoke such an inspection, may be needed. We postulate that such a "local delay" will only last a finite period of time -the requests for these inspections could be dealt with on the basis of "first come, first served"- and we shall not mention these local delays any further, because they are now irrelevant for the remainder of our considerations.)
- 2) as a result of such an inspection the person invoking it may be put

to sleep (i.e. prevented, for the time being, from proceeding with his life as prescribed by the program).

3) as a result of such an inspection, one or more sleeping persons may be woken up (i.e. induced to proceed with their life as described by the program).

We restrict ourselves to such exclusion rules for simultaneous eaters that

condition 1: if V is a permissible set of simultaneous eaters, so is any subset of V

condition 2: each person occurs in at least one permissible set of simultaneous eaters. (Note: if a person occurs in exactly one permissible set of simultaneous eaters, this set contains -on account of condition 1- only himself and no one else.)

From condition 1 it follows that there is no restriction on the set of simultaneous thinkers; as a result the inspection EXIT will never have the consequence that the person invoking it will be put to sleep. As a result, persons can only be sleeping on account of having invoked the inspection ENTRY and the act of admitting person i to the table can be associated with the waking up of person i . (A little bit more precise: if during the inspection ENTRY as invoked by person i the decision to admit him to the table is not taken, he is put to sleep, otherwise he is allowed to proceed. If in any other inspection the decision to admit person i to the table is taken (person i must be sleeping and) person i will be woken up.)

Furthermore we restrict ourselves to the case that

condition 3: for each person the action "eat" will take a finite period of time (larger than some positive lower bound and smaller than some finite upper bound); this is a contrast to the action "think" that may take an infinite period of time.

In the simplest strategy no inspection will leave a person sleeping whose admittance to the table is allowed as far as the occupancy of the table is concerned. Alas, such a strategy may have the so-called "danger of individual starvation", i.e. although all individual eating actions take only a finite period of time, a person may be kept hungry for the rest of

his days. (The classical configuration showing this phenomenon is the Problem of the Dining Quintuple. Here five places are arranged cyclically around a round table and each of five persons has his own place at the table. The restriction is that no two neighbours may be eating simultaneously. The rule will then be that every person is admitted to the table as soon as he is hungry and ~~one~~^{none} of his two neighbours is eating. In this particular example this rule leaves no choice, i.e. when I leave the table the decisions whether my lefthand neighbour and my righthand neighbour have to be admitted to the table are independent of each other. In this example my two neighbours can starve me to death, viz. when my eating lefthand neighbour never leaves the table before my righthand neighbour is eating and vice versa. If the remaining two persons remain thinking, access to the table will never be denied to my neighbours and with me hungry the process can continue for ever.) The moral of the story is that if we are looking for strategies without the danger of individual starvation, we must in general be willing to consider allocation strategies in which hungry persons will be denied access to the table in spite of the circumstance that the occupancy of the table is such that they could be admitted to the table without causing violation of the given simultaneity restrictions. Our interest in strategies without the danger of individual starvation was roused by the sobering experience that quite a few intuitive efforts to exorcize this danger led to algorithms that turned out to be quite ineffective because they could lead to deadlock situations. The remaining part of this paper deals with a general characterization of strategies that do not contain the danger of individual starvation. We shall restrict ourselves to strategies where the decision to admit persons to the table will be taken for one person at a time; from our analysis it will follow that then our characterization can be given independent of the specific simultaneity restrictions (provided of course they satisfy our stated conditions 1 and 2).

We start by proving a theorem, in which we consider the following (possible) properties of a strategy.

property A: the existence of at least one sleeping person implies at least one person who is eating or leaving the table

property B: for any person i it can be guaranteed that during a period of his hungryness the decision to admit someone else to the table will not be taken more than N_i times, where N_i is a given,

finite upper bound for person i .

Our theorem asserts that when conditions 1, 2 and 3 are fulfilled, properties A and B are the necessary and sufficient conditions for any strategy in order not to contain the danger of individual starvation.

The necessity of property A follows from the inadmissibility of the situation in which one or more persons are sleeping while all remaining ones (if any) are thinking. The thinking one may go on thinking for ever, as a result no new inspections will be evoked and the sleeping ones remain hungry for an infinite period of time.

We say that the danger of individual starvation is absent when the hungriness of any person will never last longer than a given, finite period of time. The minimum time taken by the act of eating imposes an upper bound on the personal frequency with which any given person can be admitted to the table; the total number of persons is M and therefore there is an upper bound on the total frequency with which someone is admitted to the table. Therefore the number of admissions during a period of hungriness of person i must always be less than a fixed, finite value: property B is necessary in the sense that a set of fixed, finite N_i 's exists such that it is satisfied.

Next we show that the conditions are sufficient. When person i becomes hungry -by invoking ENTRY- we have to show that his hungriness will only last a finite period of time. If in the course of that very inspection he is admitted to the table, it is true (for inspections take only a finite period of time), otherwise he ^{goes to sleep} ~~becomes sleeping~~. At the end of that inspection at least one person is eating (on account of property A) ↓ From the fact that the action "eat" takes only a finite period of time, the persons now eating will have finished doing so and will have left the table within a finite period of time. From this and property A it follows that within a finite period of time a new person will have been admitted to the table. The assumption that person i remains hungry for ever implies that the lucky person must have been someone else, i.e. within a finite period of time the number of times it has been decided during hungriness of person i that someone else is admitted to the table is increased by one. Then the argument can be repeated and within a finite period of time the number of times

someone else is admitted to the table would exceed N_i , contrary to our property B. Therefore person i will not remain hungry for ever.

Having established that properties A and B are necessary and sufficient for the absence of the danger of individual starvation, we are now in a position to characterize all strategies satisfying them with a priori given bounds N_i . For this purpose we associate with each hungry person a counter called "ac" (short for "allowance count"). Whenever person i becomes hungry, his ac is added to the set of ac's with the initial value $= N_i$; whenever it is decided that a person is admitted to the table, his ac is taken away from the set of ac's and all remaining ac's are decreased by 1. Property B is guaranteed to hold when no ac becomes negative.

We call the set of ac's "safe" when for all $k \geq 0$ holds that at most k ac's have a value $< k$. Note that also the empty set is safe. (Another formulation of safety is that it must be possible to order the ac's, if present, in such a fashion that the first $ac \geq 0$, the second $ac \geq 1$, the third $ac \geq 2$, etc.) Such a safe set has four important properties.

Property 1. No safe set contains a negative element (substitute $k = 0$ in the first definition).

Property 2. If removal of an element from the set is accompanied by a decrease by one of the remaining elements, each non-empty safe set contains at least one element that can be removed such that the remaining set is again safe: for this purpose it is sufficient -although one has often greater freedom- to choose one of the smallest values.

Property 3. Removal of any non-negative element from an unsafe set (again accompanied by a decrease by 1 of the remaining elements) will leave an unsafe set. If the set contains a negative element to start with, that one will not be removed and the remaining set remains unsafe; if the unsafe set contains no negative elements there must exist a positive k such that more than k elements have a value less than k : if the element removed is among them we are left with more than $k-1$ (≥ 0) elements with a value less than $k-1$, otherwise with even more than k elements with such a value.

Property 4. Addition of a new element will never make an unsafe set into a safe one.

Because the empty set is safe again, properties 3 and 4 tell us that an unsafe set of ac's is bound to lead to negative ac's before it is empty. Therefore safety of the set of ac's must be maintained if we wish to guarantee property B. This imposes a lower bound on the initial values of the ac's, i.e. the N_i . With M processes, at most $M-1$ will be sleeping (property A), if we impose

$$N_i \geq M - 2$$

we can guarantee that, given a safe set of ac's, the addition of a new ac will never lead to unsafety.

We can now characterize all strategies satisfying properties A and B. We call a person "admissible" if the following three conditions all hold

- 1) he must be hungry
- 2) his addition to the set V of eating persons would not cause violation of the simultaneity restrictions
- 3) his removal from the set of hungry persons would leave a safe set of ac's.

We now characterize all strategies enjoying our (necessary and sufficient) properties A and B in terms of a general permission and a specific obligation.

Each inspection ENTRY or EXIT has the general permission to decide (zero or more times) to admit an admissible person to the table. However, ~~these~~ inspections that would violate property A in the case of zero admissions have the specific obligation to admit at least one person to the table. They are the ENTRY while there are no eating persons and the EXIT in which a person is the last to leave the table while there are sleeping persons. In both cases conditions 1 and 2 and property 2 of safe sets guarantee the existence of at least one admissible person.

It is clear that any such strategy will satisfy properties A and B, it is also clear that any other strategy will have to be rejected: if the general permission is violated, an erroneous admission to the table takes place or we

end up with either deadlock or violation of property B, if the specific obligation is not fulfilled property A is violated. In this sense we have characterized all strategies satisfying properties A and B.

Concluding Remarks.

For a very specific reason the result obtained seems significant. When one is making an operating system one is faced with "absolute requirements" (of a rather logical nature) on the one hand and "desires" (not necessarily all compatible with each other) on the other. In the early design phase of the THE Multiprogramming System we had the hope that all allocation strategies could be factored in the sense that first we could produce the code that would ensure non-violation of the absolute requirements, in which then all sorts of strategic routines could be plugged in, the idea being that a change of strategic routines could influence the desirability of the systems behaviour but could never lead to violation of the absolute requirements. In the later design stages we have not been able to reach that goal: we turned up with allocation strategies for which we could prove that the absolute requirements would never be violated, but each new proposed strategy required a new proof of this fact. The origin of this failure was our inability at that time to give a constructive characterization of all possible strategies that were guaranteed to meet our absolute requirements. This paper shows that -at least in the case of the chosen absolute requirements- such a characterization can be given in terms of permissions and obligations and in such a way that it has been proved that the obligation can always be fulfilled. The question under which other circumstances such characterizations can be given is now open for investigation.

Acknowledgements.

The author wishes to express his great indebtedness to A.Martin and W.H.J. Feijen, for it is due to their persistence, in a common effort to clear up our thinking about starvation, that I was guided towards the approach described in this paper.