DRAFT EWD752-0

EWD752.html

# Distributed Arbitration

by Edsger W. Dijkstra and C. S. Scholten

## 0. Problem Description.

Generalizing the synchronization paradigm of the Dining Philosophers [1,2], we consider a finite undirected graph, whose vertices are called "philosophers" and whose edges are called "forks". A philosopher is a process consisting of a continued alternation of two states called "thinking" and "eating" respectively. A fork is a resource shared between the two philosophers it joins; it is either free or owned by one of those two philosophers. The statement "a philosopher is eating" is equivalent to the statement that that philosopher owns all the forks it shares with others. Note that for a complete graph the above corresponds to total mutual exclusion, with evidently "eating" corresponding to the critical section; the case of the graph being a pentagon corresponds to the paradigm mentioned above.

Even in the case of the complete 2-graph there is a problem, viz. the resolution of the contention for the single fork when the two philosophers happen to become "simultaneously" ready

to stop thinking. We postulate this problem of binary arbitration solved. The contention invites us to introduce after thinking and before eating an intermediate state called "ready to grab" the fork. We can now formulate more precisely our postulate:

P0 a fork is owned or no philosopher is ready to grab it .

With thinking periods of positive duration and eating periods of finite duration — and we shall assume this to be the case for the rest of our paper — we immediately derive from P0 the following corollary

C0: a philosopher ready to grab a fork owned by an eating philosopher will own that fork upon completion of the other's meal. (It is understood that a thinking philosopher owns no forks.)

It follows that the postulated solution for the complete 2-graph is free from the dangers of deadlock and of individual starvation.

We propose to solve the synchronization problem corresponding to a general graph in terms of binary arbitrations as described above. In order to give the adjective "distributed" the strongest

possible meaning, we accept the constraint that in any state a philosopher may be ready to grab <u>at most one specified</u> fork; as before, the state "ready to grab a specified fork" is terminated only by grabbing it. Consequently, a philosopher is from now on a process consisting of a continued repetition of:

thinking;
grabbing its forks in some order;
eating.

For the sake of realism we admit after the atomic act of grabbing a fork a finite delay before the new owner enters its next state, in particular before it becomes ready to grab the next fork. For our later analysis it is then irrelevant whether, at the end of an eating period, the forks are released simultaneously or not.

## 1. Avoiding the danger of deadlock.

As long as a philosopher is ready to grab a fork (which — see P0 — is owned by another) it is said to be "blocked" by that fork; the fork in question is called the "blocking fork" of that philosopher.

Deadlock - sometimes called "partial deadlock" - occurs when there exists a philosopher such that there is no possible future state in which it is eating. Without constraints on the orders in which philosophers grab their forks, the danger of deadlock is present when the graph contains a cycle: let each philosopher on that cycle grab its "left fork" first. In the sequel we shall formulate the precise constraint on the grabbing orders that exorcizes the danger of deadlock.

For each philosopher in a connected subset of (two or more) philosophers we define "its first fork in that subset" to be the first fork in its grabbing order that it shares with any other philosopher in that subset. In each connected subset "directed paths of first forks" are recursively defined by

1) a single philosopher is a directed path of first forks
2) a directed path of first forks extended with the first fork of its last philosopher is a directed path of first forks.

Since each philosopher in a connected subset has a first fork in that (finite) subset, cyclic directed paths of first forks exist.

Theorem 0. "the danger of deadlock is present" = "there exists a connected subset containing a cyclic directed path of more than two first forks". (End

of Theorem 0.)

Remark. Note that in the terminology used, a "cyclic directed path of two first forks" corresponds to a single fork that is first fork at both ends. (End of Remark.)

Proof of Theorem 0.

Assume the existence of a subset containing a cyclic directed path of three or more first forks. The state in which each philosopher on the cyclic path owns its first fork can be reached when all the philosophers outside the subset are thinking. Since none of the philosophers on the cycle can be the first to eat, deadlock occurs. Hence the truth of the right-hand side implies the truth of the left-hand side.

Assume deadlock. This implies eventually the existence of a directed cycle of more than two blocking forks, hence of a directed cycle of forks each owned by, say, its target. If there exists an owned "diagonal" fork of the cycle, there exists a smaller directed cycle of at least three owned forks (built from the "diagonal" and one part of the original cycle). By induction we conclude the existence of a directed cycle of at least three owned forks, no diagonal of which is owned. In the subset formed by the philosophers on such a cycle, the forks of the cycle are the first forks of their owners. Hence the truth of the left-hand side

implies the truth of the right-hand side. (End of Proof of Theorem 0.)

We state the following corollary of Theorem 0:

C1: in a deadlock-free system with at least one fork, there exists at least one fork that is first fork at both ends.

Remark. Since each fork blocks at most one philosopher at a time, absence of the danger of deadlock implies, thanks to P0, the absence of the danger of individual starvation. (End of Remark.)

In the sequel we confine our attention to systems free from the danger of deadlock.

## 2. Blocking paths and delays.

So-called "blocking paths" are defined recursively by

1) a single philosopher is a blocking path
2) a blocking path extended with the blocking fork of its last philosopher is a blocking path.

Let X be a philosopher. For the blocking paths that start at X we define the following meaningful

order. Consider two such paths:

1) if the one path is an extension of the other, the extended path precedes the other
2) if neither is an extension of the other, they contain a philosopher Y such that both paths are identical from X to Y, but are extended by different blocking forks of Y; the order in which Y grabs these two forks equals the order of the corresponding paths.

The above ordering of the blocking paths that start at X is meaningful for the following reason. Associate with each blocking path the following state: philosophers not on the path are thinking and the last philosopher of the path is eating. Upon completion of the current meal, a possible successor state of the system is the one associated with the next blocking path, where "next" is to be understood in terms of the above ordering of the blocking paths starting at X. (See the last paragraph of Section 0.)

Whenever X is not thinking, there is a unique blocking path starting at X and with its last philosopher eating. It now follows that the maximum number of meals that can take place between two successive thinking periods of X - called "the maximum delay for X" - equals the number of blocking paths starting at X.

From the previous paragraph it follows that a delay equal to the number of blocking paths starting at X can occur; from corollary C0 it follows that the delay for X cannot exceed that number.

## 3. Minimal average maximum delay for mutual exclusion.

For the sake of simplicity we shall confine ourselves in the following two sections to complete graphs, i.e. a fork for each pair of philosophers. This corresponds to total mutual exclusion of the eating periods. From Section 2 we conclude that the problem of minimizing maximum delays boils down to minimizing numbers of blocking paths; minimizing, for a given set of philosophers, their average maximum delay is, therefore, tantamount to minimizing the total number of blocking paths.

For N philosophers we shall now determine the minimum number of blocking paths. Let in a deadlock-free system fork f, shared by philosophers X and Y, be first fork at both ends; the existence of f is stated in corollary C1. The blocking paths of the system can now be partitioned as follows:

H blocking paths containing neither X nor Y

BX blocking paths containing X but not Y and blocking paths starting at Y with X in

second position

BY blocking paths containing Y but not X and blocking paths starting at X with Y in second position.

Note. Because f is the first fork of both X and Y, a blocking path containing both starts at one of them with the other in second position. (End of Note.)

Let us call SX the reduced system that remains when philosopher Y is removed (the forks shared by Y being removed from the grabbing orders of the others); for this reduced system, X will be denoted as "the pivotal philosopher". The system SY is similarly defined by interchanging X and Y.

A blocking path from category A occurs in both reduced systems, independently of the place a pivotal philosopher occupies in the grabbing orders of the non-pivotal ones and vice versa.

The blocking paths from category BX are fully determined by the structure of SX, and similarly for BY.

Let PX and PY be the number of blocking paths in BX and BY respectively. In a system with a minimal number of blocking paths, PX=PY

holds, because if $PX < PY$, we can decrease $PY$ without affecting either $PX$ or the number of A-paths by choosing the structure of $SY$ equal to that of $SX$. (Since deadlock implies infinitely many blocking paths, this transformation does not introduce the danger of deadlock.)

In short, heading for systems with a minimal number of blocking paths, we can confine ourselves to systems with $SX$ and $SY$ of equal structures (i.e. replacing $X$ in $SX$ by $Y$ yields $SY$).

The equal structure of $SX$ and $SY$ implies that $X$ and $Y$ have equal grabbing orders and that each non-pivotal philosopher has the forks it shares with $X$ and $Y$ in adjacent positions in its grabbing order. Thus we have to focus our attention on the reduced structure of N-1 philosophers. Repeated application ends with a structure of two philosophers.

The inverse process starts with two philosophers connected by one fork, and builds up the system by increasing the number of philosophers by one in each step. Such a step consists of "doubling" a selected philosopher. The fork connecting the pair is first fork for both, the rest of their grabbing orders is a copy of that of the selected

philosopher; the grabbing orders of the other philosophers are "stretched" at the position of the selected philosopher, whose entry is replaced by the resulting pair in some order.

The structure satisfies relation

R: the number of philosophers in the structure equals 1 or the philosophers can be partitioned into two non-empty subsets V and W such that
1) each philosopher grabs the forks it shares with members of its own set before it grabs those it shares with members of the other set, and
2) the structures corresponding to V and W both satisfy R.

This is vacuously true to start with and remains true under the doubling operation: when a selected philosopher is doubled, the resulting pair belongs to the same set(s) as the philosopher it replaces.

A blocking path starting at a philosopher from V either has all its philosophers in V or is any of those paths extended with one arbitrary philosopher from W. (Firstly, such an extension is a blocking path; secondly, it cannot be extended with a further philosopher from W, since the last

philosopher of the path has already grabbed all forks it shares with other members of W; thirdly, it cannot be extended with a philosopher from V either, since the last V-philosopher on the path owns all the forks it shares with the other members of V.) It follows that the total number of blocking paths equals:

$$(NW+1)\cdot PV + (NV+1)\cdot PW \quad ,$$

where NV = number of philosophers in V
NW = number of philosophers in W
PV = number of blocking paths in V
PW = number of blocking paths in W .

Let $f(x)$ be the minimum number of blocking paths in a system with $x$ philosophers. Since our freedom in selecting the philosopher to be doubled enables us to realize any partitioning (and subsequent subpartitionings), our previous result tells us that $f$ is the minimal solution of

for $x=1$ : $f(x)=1$
for $x>1$ : $f(x)=(n+1)\cdot f(m)+(m+1)\cdot f(n)$
where $m\geq 1$, $n\geq 1$, $m+n=x$

It can be shown that the minimal solution is obtained by choosing $|m-n|\leq 1$. (See Appendix.)

## 4. Configurations symmetric in the philosophers.

From the previous section it follows that the minimal number of blocking paths can always be realized by assigning to each fork a natural number, called its "rank", such that

1) forks meeting at a philosopher have distinct ranks, and
2) each philosopher grabs its forks in the order of increasing rank.

We call such systems "ranked systems". (Ranked systems are obviously deadlock-free.)

Example. By way of illustration we give a system that is deadlock-free, is symmetric in the philosophers, and realizes the minimum number of blocking paths, but cannot be ranked. With four philosophers numbered 0,1,2,3 and addition being modulo 4, philosopher $i$ grabs the forks it shares with $i+2$, $i+1$, and $i+3$ in that order. We don't need to consider that system; there is, indeed, a ranked system enjoying the same properties: it suffices to re-interpret "addition" as the bit-wise sum modulo 2. (End of Example.)

In general, for a ranked system of N philosophers, the number of distinct ranks exceeds the number of forks meeting at a philosopher, e.g. N=3 requires 3 distinct ranks. In such a case,

symmetry in the philosophers is precluded: in a symmetric ranked system each philosopher grabs a fork of each rank in the system.

In a symmetric ranked system, forks of two distinct ranks form cycles of length 4; their diagonals have a third rank.

Proof. Let p and q be two distinct ranks. Consider the two 2-fork paths (p,q) and (q,p), starting at the same philosopher. The latter shares with their endpoints forks which, for reasons of symmetry, have the same rank, r say. Hence their endpoints coincide. Obviously, the other diagonal is of rank r as well. (End of Proof.)

Consider a symmetric ranked system of more than 2 philosophers. Two ranks partition therefore the philosophers into partitions of 4. Forks of a next rank either connect philosophers from the same partition, or pair partitions, thereby doubling their size and halving their number. The number of philosophers in a symmetric ranked system is, therefore, a power of 2.

Furthermore a symmetric ranked system exists for each $N = 2^k$. Number the philosophers (in binary) from 0 through N-1, and assign to the fork shared by philosophers i and j the

rank $i+j$, where "addition" is again the bit-wise sum modulo 2. If X is the number of a philosopher, we can renumber the philosophers without affecting the ranks by performing for each philosopher number $i$: $i := i + X$ (addition again bit-wise modulo 2). In the new system — which is congruent with the old one — philosopher X has become philosopher 0, and thus the symmetry has been established.

It follows from our previous section that for $N=2^k$ the above symmetric ranked system realizes the minimum number of blocking paths, hence the minimum maximum delay for each philosopher. It follows from that same section that this delay equals unfortunately

$$(\text{PROD } i : 0 \leq i < k : 1+2^i) ,$$

i.e.

| k: | delay: | N: | |
|---|---|---|---|
| 0 | 1 | 1 | |
| 1 | 2 | 2 | |
| 2 | 6 | 4 | |
| 3 | 30 | 8 | |
| 4 | 270 | 16 | and so on. |

The exploding worst case casts some doubts on the general utility of binary arbitration as a means for implementing mutual exclusion. We can only recommend it when it can be shown on other

grounds – such as timing considerations – that the worst case delay will not occur.

## 5. A two-stage solution.

The case of the complete graph, corresponding to total mutual exclusion, was studied in order to determine how bad worst-case delays could be. We now return to the general graph, for which we shall propose a different solution. At the level of detail in which it will be described, it uses less austere communication facilities than binary arbitration only, but it achieves the smallest possible worst-case delay.

For this purpose, the life of a philosopher is viewed as a cyclic succession of three states, called "thinking", "hungry", and "eating" respectively; for brevity's sake the union of the last two states will be denoted by "tabled".

Furthermore, also each fork has three states: either it is undirected or it is directed, i.e. it carries an arrow in one of the two possible directions.

The life of a philosopher is now programmed as follows, angle brackets being used to delimit

"point actions"; philosophers sharing a fork are called each other's "neighbours".

```
do true →
        thinking;
T0:     < directs an arrow towards each of its tabled neigh-
          bours and switches to hungry, hence to tabled >
        hungry;
T1:     < observes to be without outgoing arrows and
          switches to eating, hence remains tabled >
        eating
T2:     < makes all its forks undirected and switches
          to thinking >
od
```

The system, which is started with all philosophers thinking and all forks undirected, maintains the following invariants:

Q0: for each pair of neighbours X and Y
"both X and Y are tabled" =
"the fork shared by X and Y is directed"

Q1: for each philosopher X
"X is hungry" or "X is without outgoing arrows".

As a consequence of Q0, a thinking philosopher has all its forks undirected; as a consequence of Q1, T2 removes only incoming arrows. The universal truth

truth of Q0 and Q1 precludes simultaneously eating neighbours.

In order to prove that transitions T1 – which are implemented by making a philosopher "wait" until all its outgoing arrows have disappeared – cannot lead to deadlock, we observe the obvious invariance of

Q2 : directed forks do not form directed cycles.

The system is free from the danger of individual starvation. Observe a hungry philosopher X and all philosophers reachable from X via directed forks. All philosophers of this directed subgraph are tabled; hence the subgraph cannot grow. It is furthermore cycle-free (on account of Q2): its longest path ends at a philosopher that can eat and, hence, that maximum path length gives an upper bound for the delay of X. With N philosophers, the delay is therefore at most N.

Nice as this solution is, it begs the question, since the only safe way of implementing the T's – in particular T0 and T2 – as sequential processes that we know of is to make the T's of neighbours mutually exclusive in time, and that mutual exclusion was the original problem.

Under the assumption that both thinking and eating of a philosopher always require more than N "transition times T", an implementation of the mutual exclusion of neighbouring T's by means of binary arbitration yields, on the microscopic level, a delay of at most N transition times.

The moral of the story is that an algorithm with very poor worst-case behaviour can be safely used when embedded in an environment that prevents the realization of the poor performance.

## References

[1] Hoare, C.A.R., Towards a Theory of Parallel Programming, Operating Systems Techniques, Ed. C.A.R. Hoare and R.H. Perrott, Academic Press, London and New York, 1972, pp. 61-71.

[2] Dykstra, Edsger W., Hierarchical Ordering of Sequential Processes, ibid. pp. 72-93.

## Appendix

The claim that the system

for $x = 1$: $f(x) = 1$

for $x > 1$: $f(x) = (n+1) \cdot f(m) + (m+1) \cdot f(n)$

where $m \geq 1, n \geq 1, m+n = x$

yields its minimal solution for $|m-n| \leq 1$ is equivalent to the statement that the function $f$ defined by

$$\left.\begin{aligned} f(1) &= 1 \\ f(2 \cdot k) &= (k+1) \cdot f(k) + (k+1) \cdot f(k) \\ f(2 \cdot k+1) &= (k+1) \cdot f(k+1) + (k+2) \cdot f(k) \end{aligned}\right\} \text{for } k \geq 1$$

satisfies for $m \geq 1, n \geq 1$, $P(m,n) \geq 0$ with $P$ defined by

$$P(m,n) = (m+1)\cdot f(n) - f(m+n) + (n+1)\cdot f(m) \qquad (0).$$

With Q defined by

$$Q(m,n) = m\cdot f(n) - n\cdot f(m) \qquad (1)$$

we first prove $m \leq n \Rightarrow Q(m,n) \geq 0$ (2).

Proof of (2). On account of (1), (2) follows from

$f(x)/x$ is an ascending function for $x \geq 1$ (3)

which follows from

$$n\cdot f(n+1) - (n+1)\cdot f(n) > 0 \quad \text{for } n \geq 1 \qquad (4)$$

We shall demonstrate (4) by

a) observing its truth for $n=1$ (note that $f(2)=4$),

b) demonstrating that the first difference of the left-hand side of (4) is positive. This first difference is

$$(n+1)\cdot f(n+2) - 2\cdot(n+1)\cdot f(n+1) + (n+1)\cdot f(n) =$$
$$(n+1)\cdot(f(n+2) - 2\cdot f(n+1) + f(n)) \quad \text{for } n \geq 1.$$

We are left with the proof that f has a positive second difference. With ddf defined by

$$ddf(n) = f(n+2) - 2\cdot f(n+1) + f(n)$$

we deduce from the definition of f

$$ddf(2\cdot k) = 2\cdot(f(k+1) - f(k))$$
$$ddf(2\cdot k+1) = (k+2)\cdot ddf(k) \quad .$$

Since $f(1)=1$, $f(2)=4$, $f(3)=11$ we find for $k=1$: $f(k+1)-f(k)=3$ and $ddf(k)=4$. From that base it follows by induction that all first and second differences of $f$ are positive. Hence, (2) has been established. (End of Proof of (2).)

Next, we define — with $(2\cdot k)\,\underline{\text{div}}\,2 = (2\cdot k+1)\,\underline{\text{div}}\,2 = k$ —

$$x_0 = m \;\underline{\text{div}}\; 2 \qquad x_2 = n \;\underline{\text{div}}\; 2$$
$$x_1 = (m+1)\;\underline{\text{div}}\; 2 \qquad x_3 = (n+1)\;\underline{\text{div}}\; 2 \quad .$$

These functions satisfy — as is easily verified —

$$x_0 + x_1 = m \qquad x_2 + x_3 = n$$
$$\{x_0 + x_3,\ x_1 + x_2\} = \{(m+n)\;\underline{\text{div}}\;2,\ (m+n+1)\;\underline{\text{div}}\;2\} \quad .$$

Using in the subscripts "$\dot{+}$" to denote the bit-wise sum modulo 2, we now consider the expression: (5)

$$(\underline{S}\, i : 0 \le i < 2 : (x_{i+1} + x_{i+2} + 1)\cdot P(x_i, x_{i+3}) + (x_{i+3} - x_{i+1})\cdot Q(x_i, x_{i+2}))$$

In (5) we have two types of terms

a) terms derived from the middle term of $P$; they add up to

$$-(x_1+x_2+1)\cdot f(x_0+x_3) - (x_0+x_3+1)\cdot f(x_1+x_2) =$$
$$-f(x_0+x_1+x_2+x_3) =$$
$$-f(m+n) \quad .$$

b) terms with $f(x_i)$ for $0 \leq i < 4$. The coefficient of $f(x_i)$ is

$$(x_{i+1} + x_{i+2} + 1) \cdot (x_{i+3} + 1) + (x_{i+1} - x_{i+3}) \cdot x_{i+2} =$$
$$(x_{i+2} + x_{i+3} + 1) \cdot (x_{i+1} + 1) .$$

Combination of the terms with $i=0$ and $i=1$ gives

$$(x_2 + x_3 + 1) \cdot f(x_0 + x_1) = (n+1) \cdot f(m) ,$$

combination of the terms with $i=2$ and $i=3$ gives

$$(x_0 + x_1 + 1) \cdot f(x_2 + x_3) = (m+1) \cdot f(n) .$$

Consequently, expression (5) equals $P(m,n)$.

Because $(x_{i+3} > x_{i+1}) \Rightarrow (x_{i+2} \geq x_i)$ - and, similarly $(x_{i+3} \leq x_{i+1}) \Rightarrow (x_{i+2} \leq x_i)$ - we conclude on account of (2) that the second terms in (5) are $\geq 0$. Therefore, $P(m,n) \geq 0$ can be concluded from

$P(m \text{ div } 2, (n+1) \text{ div } 2) \geq 0$ and $P((m+1) \text{ div } 2, n \text{ div } 2) \geq 0$.

This observation concludes the inductive argument.

C. S. Scholten
Philips Research Laboratories
5600 MD EINDHOVEN
The Netherlands

Edsger W. Dijkstra
Burroughs
Plataanstraat 5
5671 AL NUENEN
The Netherlands