# UNIVERSITY OF LONDON
## IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

## EXAMINATIONS 1998

MEng Honours Degrees in Computing Part IV
MEng Honours Degree in Information Systems Engineering Part IV
MSci Honours Degree in Mathematics and Computer Science Part IV
MSc Degree in Advanced Computing
for Internal Students of the Imperial College of Science, Technology and Medicine

*This paper is also taken for the relevant examinations for the*
*Diploma of Membership of Imperial College*
*Associateship of the City and Guilds of London Institute*
*Associateship of the Royal College of Science*

PAPER 4.37 / I4.12

# DISTRIBUTED ALGORITHMS

Monday, May 11th 1998, 10.00 - 12.00

*Answer THREE questions*

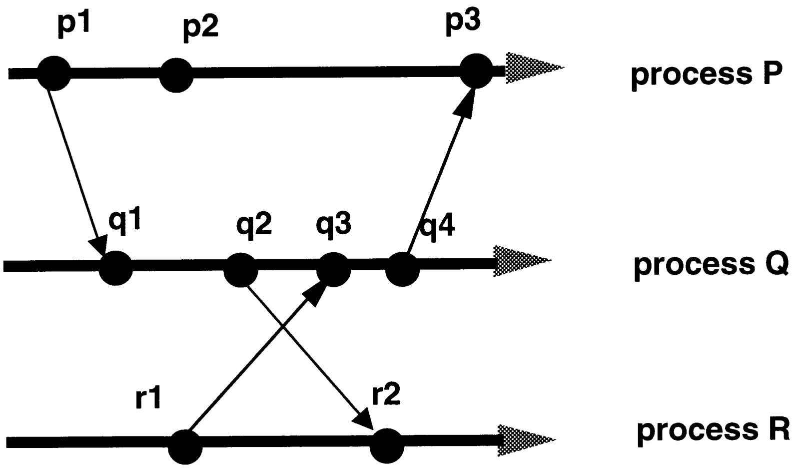
For admin. only: paper contains 4 questions

1a i) Given any two events a,b in a distributed system, define what is meant by the causal order relation a ➔ b, i.e. "a precedes b".

ii) Describe how Lamport's logical clocks can be used to provide partial ordering consistent with causal ordering, i.e. for logical timestamps C and events a and b, a ➔ b implies that C(a) < C(b).

iii) Explain why C(a) < C(b) does not necessarily imply that a ➔ b.

iv) Starting from C=0 annotate the process interaction diagram below with logical timestamp values for each event in each process P, Q and R.




b i) Describe how vector clocks can be used to provide partial ordering which reflects causal ordering precisely, i.e

a ➔ b iff VT(a) < VT(b) for vector timestamps VT and events a and b.

ii) Starting from VT=(0,0,0) annotate the process interaction diagram above with vector timestamp values for each event in each process P,Q and R.

iii) It is asserted that, at any instant,

$\forall i, \forall j$ : VT[i] in process i $\leq$ VT[i] in process j, for all i≤j.

Briefly comment on the truth of this assertion, justifying your answer.

*The two parts carry, respectively, 50%, 50% of the marks.*

© University of London 1998

2a. Consider a system consisting of a number of client processes P which need to gain access to a number of resources. *Briefly* describe the conditions which are necessary to give rise to a *resource deadlock.*

b. A distributed data base system is accessed using transactions Ti. Serializability is ensured through the use of two phase locking. However deadlocks can then occur and must be prevented. If a logical timestamp is allocated to each transaction Ti, then timestamps can be used to consistently resolve conflicting transactions locally at each resource allocator, thereby preventing deadlock.

Describe the 'Wait-Die" and 'Wound-Wait' methods for conflict resolution (suggested by Rozenkrantz, Stearns and Lewis).

*Briefly* justify that each method ensures that deadlock will not occur, and identify the method as preemptive or non-preemptive.

c. Consider an environment in which there are a number of communicating processes P. Processes can be blocked waiting for message transmission to or message receipt from other processes. Describe the conditions that give rise to a *communications deadlock.*

d. What is the main difference between the conditions for a communications deadlock and those for a resource deadlock.

*Briefly* indicate how this difference is reflected in the two probe diffusing algorithms for resource and communications deadlock detection of Chandy, Misra and Haas.

*The four parts carry, respectively, 10%, 50%, 10%, 30% of the marks.*

3a Define the Validity, Agreement and Integrity properties required of a *reliable broadcast* algorithm in an asynchronous distributed system. Define the additional property that must be satisfied by a reliable broadcast algorithm that provides FIFO delivery order.

b Explain the difference between *Causal* and FIFO broadcast. Outline a simple algorithm which uses FIFO broadcast to implement Causal Broadcast.

c Describe briefly how the ISIS CBCAST protocol is implemented and comment on why it is more efficient than the protocol you described in part b.

d Andrew, Bob and Cameron (*A*,*B* & *C*) all subscribe to an Internet mailing list on interesting car registration plates. Andrew broadcasts a request for plates beginning with *m* and Bob responds with his latest sightings *s*. However, when Cameron reads his email, he sees Bob's response (*s*) before Andrew's request (*m*) arrives. Using message diagrams, explain the operation of the protocols you described in parts b & c in resolving this situation.

*The four parts carry respectively 20%, 30%, 30%, 20% of the marks.*

*Turn over ...*

4a What are the three properties of distributed systems said to be *asynchronous*?

b State the three message passing assumptions which are made in arriving at a solution to the Byzantine Generals problems. Briefly describe what each of these assumptions implies in the construction of reliable computing systems. Explain which of the assumptions means these systems are *not* asynchronous.

c If a multicast is said to be fully *stable*, what does this mean? The following diagram is a Psync context graph in which ***ai*** is the ***i*** th message from process ***a***..

$$
\begin{array}{ccccc}
 & & a_1 & & \\
 & & \downarrow & & \\
 & & b_1 & & \\
 & \swarrow & & \searrow & \\
 & c_1 & & & a_2 \\
\swarrow & & \searrow & \swarrow & \\
a_3 & & b_2 & & \\
 & \searrow & \swarrow & & \\
 & & c_2 & &
\end{array}
$$

Explain how stability can be determined from a Psync context graph and give the general condition for a message to be stable. List the stable messages in the above context graph justifying why each is stable.

d Explain how stability is used to implement totally ordered multicasts in Psync.

*The four parts carry, respectively, 10%, 20%, 40%, 30% of the marks.*

*End of paper*

© University of London 1998