COMPUTER SCIENCE TRIPOS Part II

Tuesday 1 June 2004 1.30 to 4.30

Paper 7

Answer **five** *questions.*
Submit the answers in five **separate** *bundles, each with its own cover sheet. On each cover sheet, write the numbers of* **all** *attempted questions, and circle the number of the question attached.*

You may not start to read the questions printed on the subsequent pages of this question paper until instructed that you may do so by the Invigilator

1 Comparative Architectures

The IA-64 architecture includes a number of instruction set features that are intended to allow compilers to generate code that will execute with fewer pipeline stalls.

Explain how *four* of the following five features operate, and how they could be used by a compiler to improve performance relative to more conventional instruction sequences.

(*a*) Register windows for procedure call linkage.

(*b*) Not a Thing (NaT) "poison" bits associated with registers.

(*c*) Address aliasing table (ALAT) for load data speculation.

(*d*) Register file rotation.

(*e*) Parallel compares.

[5 marks each]

2 Digital Communication II

(*a*) Describe the operation of the TCP congestion control scheme, covering the two phases of *slow start*, *congestion avoidance* and the triggers of *duplicate ack* and *timeout* that set it in train. [10 marks]

(*b*) A computer is connected to the Internet via a wireless link using the packet service from the mobile phone company. This offers 30 kbps data rate, but has a widely varying round-trip time across the wireless hop from the computer to the next hop in the backbone of between 700 milliseconds and 1.1 seconds. It also has an average packet loss rate of 25%. Using time-sequence diagrams to illustrate the fate of data packets and acknowledgements, show the impact that this has on the throughput that a typical TCP implementation would achieve. [10 marks]

3 Security

(*a*) Describe *four* of the problems from which classical multilevel-secure systems suffer. [12 marks]

(*b*) Your client is proposing to implement an e-mail system with the property that every e-mail sent internally within the company – that is, with no outside recipients – should be deleted after 180 days unless a manager authorises its retention.

Which of the problems in part (*a*) would you have to consider, and why? [8 marks]

4 Advanced Graphics

(a) Specify an appropriate knot vector for each of the following NURBS curves.

(i) A uniform cubic NURBS curve defined by six control points.

(ii) Similar to (i) but passing through both endpoints.

(iii) Similar to (i) but passing through the third control point, possibly with lower continuity at that point.

(iv) The cubic Bezier curve defined by four control points. [8 marks]

(b) Give the continuity class for each of:

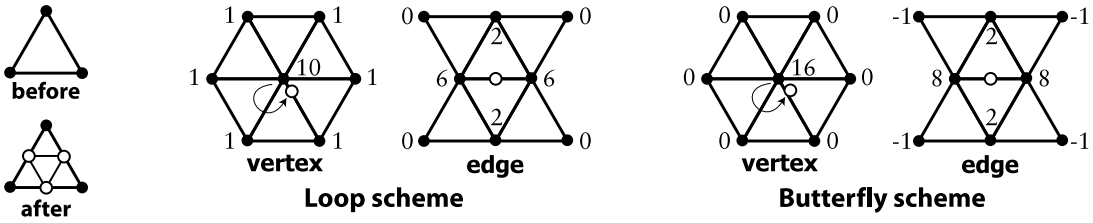
(i) curve (a)(i) *between* the knots;

(ii) curve (a)(ii) *at* the knots;

(iii) curve (a)(iii) at the third control point.

[The continuity class is the highest derivative which is guaranteed to be continuous at the point(s) in question.] [3 marks]

(c) The Loop and Butterfly subdivision schemes can both operate on triangular meshes, in which all of the polygons have three sides. Both schemes subdivide the mesh by introducing new vertices at the midpoints of edges, splitting every original triangle into four smaller triangles, as shown below. Each scheme has rules for calculating the locations of the new "edge" and "vertex" vertices based on the locations of the old vertices. These rules are shown below. All weights should be multiplied by $\frac{1}{16}$.



(i) Which of the two schemes produces a limit surface which interpolates the original data points?

(ii) Which of the four rules must be modified when there is an extraordinary vertex? For each of the four rules either explain why it must be modified or explain why it does not need to be modified.

(iii) Suggest appropriate modifications where necessary to accommodate extraordinary vertices.

[9 marks]

5 Computer Systems Modelling

(*a*) Consider an interactive system with average system response time R and system throughput X. Derive the *interactive response time law* when there are M users present, each with an average think time Z. [6 marks]

(*b*) Consider an interactive timesharing system with M users, one CPU and two disks labelled A and B respectively. Suppose that measurements of the system have determined that: the average user think time Z is 5 seconds; the average time spent by a job in the CPU is 40 ms per request; the average time spent by a job at the individual disks per request is 30 ms for disk A and 25 ms for disk B; the visit counts per job are 25 requests for the CPU, 20 requests for disk A and 4 requests for disk B.

(*i*) For each of the devices: CPU, disk A and disk B, determine the service demands. [4 marks]

(*ii*) If disk A has a utilisation of 60%, what is the utilisation of the CPU and of disk B? [5 marks]

(*iii*) If the utilisation of disk B is 10%, what is the average response time when there are 20 users present? [5 marks]

6 Specification and Verification I

Write short notes (approximately one paragraph) explaining the similarities and differences between each of the following:

(*a*) Annotations and verification conditions. [5 marks]

(*b*) Partial and total correctness. [5 marks]

(*c*) Deep and shallow embedding. [5 marks]

(*d*) First order and higher order logic. [5 marks]

7 Specification and Verification II

(*a*) Explain how combinational and sequential devices can be modelled in higher order logic in a uniform way (i.e. so that combinational and sequential devices can be connected). [3 marks]

(*b*) Illustrate your explanation by showing how to define combinational devices NOT and AND that perform negation and conjunction, respectively, and a sequential unit-delay device DEL such that the output of DEL at time $t{+}1$ is the value input at t. [3 marks]

(*c*) Define in higher order logic a predicate Rose such that if $t > 0$ then Rose f t is true if and only if f has a rising edge at time t (i.e. f is true at t but false at $t{-}1$). [2 marks]

(*d*) Draw a diagram showing how to connect instances of NOT, AND and DEL to implement a device RoseImp, with one input and one output, such that the output is true at time t, where $t > 0$, if and only if there is a rising edge on the input at time t. [6 marks]

(*e*) Represent your diagram in higher order logic by defining a predicate RoseImp, and then outline how to show that:

$$\forall in\ out.\ \mathsf{RoseImp}(in, out)\ \Rightarrow\ \forall t.\ out(t{+}1) = \mathsf{Rose}\ in\ (t{+}1)$$

You need not give a detailed proof, just an overview of how such a proof could be performed. [6 marks]

8 Information Theory and Coding

(a) Consider an alphabet of 8 symbols whose probabilities are as follows:

A	B	C	D	E	F	G	H
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{128}$	$\frac{1}{128}$

(i) If someone has selected one of these symbols and you need to discover which symbol it is by asking "yes/no" questions that will be truthfully answered, what would be the most efficient sequence of such questions that you could ask in order to discover the selected symbol? [2 marks]

(ii) By what principle can you claim that each of your proposed questions is maximally informative? [2 marks]

(iii) On average, how many such questions will need to be asked before the selected symbol is discovered? [2 marks]

(iv) What is the entropy of the above symbol set? [2 marks]

(v) Construct a uniquely decodable prefix code for the symbol set, and explain why it is uniquely decodable and why it has the prefix property. [2 marks]

(vi) Relate the bits in your prefix code to the "yes/no" questions that you proposed in (i). [2 marks]

(b) Explain the meaning of "self-Fourier", and cite at least *two* examples of mathematical objects having this property. [3 marks]

(c) Explain briefly:

(i) sensation limit; [1 mark]

(ii) critical band; [1 mark]

(iii) Bark scale. [1 mark]

(d) Which different aspects of perception do Weber's law and Steven's law model? [2 marks]

9 Quantum Computing

(*a*) You are given a qubit that is in one of two states: either $|\phi\rangle = |0\rangle$ or $|\psi\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle$. If you measure the qubit in the computational basis, what is the probability that your measurement correctly identifies the state? [3 marks]

(*b*) Draw a labelled schematic circuit diagram for:

(*i*) the phase estimation algorithm; [4 marks]

(*ii*) Grover's algorithm. [4 marks]

(*c*) Suppose a search problem has an unknown number M of marked states. Show how phase estimation and Grover's algorithm can be combined to estimate M to a high accuracy using $O(\sqrt{N})$ oracle calls. [Hint: The Grover iterate, G, has eigenvalues $e^{\pm i\theta}$ where $\sin^2(\theta/2) = M/N$.] [5 marks]

(*d*) Suppose there is an algorithm which can determine the number M of marked states in an unsorted search space of size N using only $O(\log(N))$ oracle calls. Explain why this would allow us to solve NP-complete problems in polynomial time. [4 marks]

10 Digital Signal Processing

(*a*) You have designed a digital water-level display installed on the River Cam. A sensor measures the current height of a small floating ball once every minute. In order to reduce the fluctuations that small waves would otherwise cause in the displayed value, you implemented a digital filter $y_i = 0.8y_{i-1} + 0.2x_i$, where the x_i are the measured and the y_i are the displayed water levels.

(*i*) What type of filter is this? [2 marks]

(*ii*) The standard deviation caused by small waves in the measurements is 30 mm. There is no measurable correlation between these added noise values. Calculate the standard deviation caused by small waves in the displayed water levels. [8 marks]

(*b*) Let H be a digital low-pass filter with finite impulse response $h_0, h_1, \ldots, h_7$. Let f_s be the sampling frequency. Give the impulse response $h'_0, h'_1, \ldots, h'_7$ of a filter H' with frequency response $|H'(f)| = |H(f_\mathrm{s}/2 - f)|$. [4 marks]

(*c*) A programmer cuts a block out of a digitised sound signal and applies the Discrete Fourier Transform to estimate its spectral power distribution.

(*i*) What effect distorts the resulting power spectrum? [3 marks]

(*ii*) Describe briefly *one* technique to reduce these distortions. [3 marks]

11 Topics in Concurrency

A *simulation* between CCS terms is defined to be a binary relation S between CCS terms such that whenever $(t, u) \in S$ for all actions a and terms t'

$$t \xrightarrow{a} t' \Rightarrow \exists u'.\ u \xrightarrow{a} u' \ \&\ (t', u') \in S\ .$$

Write $t \leq u$ iff there is a simulation S for which $(t, u) \in S$.

(*a*) Write down the transition rules for CCS parallel composition. Show that for CCS terms if $t \leq u$ and $t' \leq u'$, then $(t \parallel t') \leq (u \parallel u')$ holds between their parallel compositions. [6 marks]

(*b*) By exhibiting suitable CCS terms show that $t \leq u$ and $u \leq t$ together does not necessarily imply that t and u are strongly bisimilar. [6 marks]

(*c*) Consider the following fragment of Hennessy–Milner logic:

$$A ::= \langle a \rangle A \mid \bigwedge_{i \in I} A_i\ ,$$

where a is an action of CCS and I is a set. Show that $t \leq u$ iff for all assertions A in the fragment, whenever t satisfies A then so does u. [8 marks]

12 Information Retrieval

(*a*) What role does stemming play in automatic indexing? [4 marks]

(*b*) Briefly describe the principles behind the Porter Stemmer. [5 marks]

(*c*) One extreme form of "stemming" is the mapping of all words of a certain prefix length onto one term. What effect will this prefix mixing have if the prefix length is extremely short, for example only two letters long? Compare this with a situation with more linguistically motivated stemming. [5 marks]

(*d*) Consider a query with four relevant documents, and a ranked IR system which returns them in the following order ("X" represents a relevant document, "–" represents an irrelevant document).

Rank	Relevance
1	X
2	X
3	–
4	–
5	–
6	–
7	–
8	–
9	–
10	X
11	X

Calculate this system's mean precision at seen documents and 11-point average precision. [6 marks]

13 Business Studies

A software project has two phases, each with three tasks. They are expected to take the following amount of effort:

Phase 1:	
Analysis:	3 weeks
Code:	2 weeks
Test:	1 week
Phase 2:	
Analysis:	1 week
Code:	2 weeks
Test:	3 weeks

Within a phase a task cannot start until the previous task completes. A task in Phase 2 cannot start until the corresponding task in Phase 1 has completed.

(*a*) Draw the PERT and Gantt charts for the project. What is the minimum elapsed time? [5 marks]

(*b*) Two staff are assigned to the project, an analyst and a programmer. The analyst also acts as test engineer. How long will the project take with this staffing? [5 marks]

(*c*) The analyst is charged out at a fixed rate of £500/day and the programmer at £300/day, including overheads. How much would you quote for the project and why? [5 marks]

(*d*) Explain how you would monitor such a project. How would you turn the result into a product? [5 marks]

14 E-Commerce

A company proposes to introduce a new e-mail service, one with a small postage charge for each message. It is believed that by charging for each message sent the amount of spam will be reduced.

The response rate for spam is about 0.005%, that is about five in a hundred thousand messages elicit a positive response.

(*a*) Estimate a price that might economically discourage spam. [5 marks]

(*b*) Describe approaches to handling the micro-payments involved. [5 marks]

(*c*) What is meant by a *network externality*, and how might it be overcome in this case? [5 marks]

(*d*) The postage charges might be credited to the recipient, less a handling fee. Describe a mechanism for this. [5 marks]

15 Additional Topics

VNC is an ultra-thin client system based on a simple display protocol that is platform-independent.

(*a*) Describe the advantages of a stateless architecture as used in VNC. [2 marks]

(*b*) List the basic protocol interactions. [6 marks]

(*c*) Give *four* pixel data encodings. [6 marks]

(*d*) Comment on the tradeoff between compression factor and complexity. [6 marks]

16 Additional Topics

Guy Fawkes and TESLA are streaming authentication protocols based on hash chains.

(*a*) Describe the problem that Guy Fawkes was designed to solve. [2 marks]

(*b*) Briefly describe how Guy Fawkes works. Draw a diagram of a Guy Fawkes packet (or several) and explain the function and purpose of all the parts. [8 marks]

(*c*) List and explain the practical shortcomings of Guy Fawkes. [4 marks]

(*d*) Draw a diagram of a chain of TESLA packets. For each of the Guy Fawkes shortcomings you found in part (*c*), explain how TESLA fixes it, if it does. [6 marks]

END OF PAPER