COMPUTER SCIENCE TRIPOS Part II

Tuesday 7 June 2005 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

STATIONERY REQUIREMENTS
Script Paper
Blue Coversheets
Tags

1 Comparative Architectures

(*a*) Most RISC architectures use a 32-bit fixed length instruction encoding. In contrast, Intel x86 and VAX use a variable length encoding, and the IA-64 uses 128-bit instruction "bundles". Compare and contrast these different instruction encodings, with particular reference to their ease (or otherwise) to decode in a super-scalar implementation, and the "code density" they achieve. [10 marks]

(*b*) Discuss the pros and cons of architectures with a 64-bit word size *versus* those with a 32-bit word size. Which applications are likely to benefit most? [4 marks]

(*c*) If you were a processor architect targeting embedded applications where memory is a scarce resource, how might you design a RISC-like instruction set that will achieve efficient use of memory? [6 marks]

2 Digital Communication II

(*a*) Explain the rôle of the checksum mechanisms in OSI layers 2, 3 and 4 of the Internet Protocols. [9 marks]

(*b*) An ARM assembly code implementation of the IP/TCP checksum function appears below. Annotate this with comments that explain the code.

```
        sub     r3, r3, #16
loop:
        ldmia   r2!, {r6-r9}
        adds    r4, r4, r6
        adcs    r4, r4, r7
        adcs    r4, r4, r8
        adcs    r4, r4, r9
        adcs    r4, r4, #0
        adc     r4, r4, #0
        subs    r3, r3, #16
        bxx     loop
        add     r3, r3, #16
```

[6 marks]

(*c*) *traceroute* is a commonly used diagnostic tool in the Internet. Explain how it works, making reference to its use of the *TTL Expiry* facility of *ICMP*, and the circumstances under which it might generate misleading or incorrect output. [5 marks]

3 Advanced Graphics

Brian and Geoff Wyvill developed a blobby object modelling method where the blobby object is defined by a number, n, of centres, $\mathbf{P}_i$, each with an associated radius, R_i. They define a function

$$g(r, R) = \begin{cases} 1 - \frac{4}{9}\frac{r^6}{R^6} + \frac{17}{9}\frac{r^4}{R^4} - \frac{22}{9}\frac{r^2}{R^2}, & r \leq R \\ 0, & r > R \end{cases} \tag{1}$$

and sum the contributions from all centres to give a function over all space

$$F(\mathbf{P}) = \sum_{i=1}^{n} g(|\mathbf{P} - \mathbf{P}_i|, R_i) \tag{2}$$

The surface of the blobby object is defined as all points, $\mathbf{P}$, where

$$F(\mathbf{P}) = \tfrac{1}{2}. \tag{3}$$

(*a*) Sketch, in 2D, the 2D blobby "surface" for each of the following cases:

(*i*) $n = 2$, $\mathbf{P}_1 = (0, 0)$, $R_1 = 2$, $\mathbf{P}_2 = (4, 0)$, $R_2 = 2$;

(*ii*) $n = 2$, $\mathbf{P}_1 = (0, 0)$, $R_1 = 2$, $\mathbf{P}_2 = (2, 0)$, $R_2 = 2$;

(*iii*) $n = 2$, $\mathbf{P}_1 = (0, 0)$, $R_1 = 2$, $\mathbf{P}_2 = (3, 0)$, $R_2 = 4$. [6 marks]

(*b*) Outline an algorithm which will generate a reasonable approximation, in 3D, to the 3D blobby surface (equation 3) which could be drawn by a graphics card that can draw only triangles. [10 marks]

(*c*) Describe variations of equation 2 which allow for:

(*i*) CSG union of blobby objects;

(*ii*) CSG intersection of blobby objects;

(*iii*) CSG difference of blobby objects. [4 marks]

4 Security

(*a*) Describe briefly security policy models that might be suitable for protecting

(*i*) medical records; [4 marks]

(*ii*) police intelligence data; [4 marks]

(*iii*) school records. [4 marks]

(*b*) The Children Act 2004 empowers the Government to establish child protection databases for England and Wales which will, it is hoped, identify cases of child abuse at an early stage. These databases will be fed with medical and school records, police intelligence data, and social work assessments. Social workers, doctors, nurses, teachers and police officers will be able to query them. Sketch a possible security policy for such a database, and discuss the most likely implementation problems. [8 marks]

5 Computer Systems Modelling

Consider the s server $M/M/s$ queue with arrival rate λ and service rate μ.

(*a*) Consider the birth–death process modelling the number of customers present and sketch the state space diagram showing the states of the system, the possible transitions between the states and the transition rates. [4 marks]

(*b*) Write down the detailed balance equations for this birth–death process. [4 marks]

(*c*) Find the equilibrium distribution, π_n $(n = 0, 1, \ldots)$, for the number of customers present by solving the detailed balance equations in terms of π_0. [6 marks]

(*d*) By normalising these expressions for π_n, find an expression for π_0 and hence show that the equilibrium distribution exists if and only if $\lambda < s\mu$. [6 marks]

6 Specification and Verification I

(*a*) State and explain Hoare's assignment axiom for simple assignments $V{:=}E$, where V is a variable. [5 marks]

(*b*) Is Hoare's assignment axiom valid for assignments $V := E$ if the expression E can have side effects? Justify your answer. [5 marks]

(*c*) State and explain the assignment axiom for array assignments $V(E_1) := E_2$, where V is an array variable. [5 marks]

(*d*) The following alternative "forward" rule for assignments has been proposed:

$$\vdash \{P\}\ V{:=}E\ \{\exists v.\ V = E[v/V]\ \wedge\ P[v/V]\}$$

Explain informally why this rule is valid. [5 marks]

7 Specification and Verification II

(*a*) What is *model checking*? [4 marks]

(*b*) Distinguish between *explicit state* and *symbolic* model checking. [4 marks]

(*c*) Describe briefly the *state explosion problem* in formal verification. [4 marks]

(*d*) How might theorem proving help with the state explosion problem? [4 marks]

(*e*) How might abstraction help with the state explosion problem? [4 marks]

8 Information Theory and Coding

(*a*) For a binary symmetric communication channel whose input source is the alphabet $X = \{0, 1\}$ with probabilities $\{0.5, 0.5\}$ and whose output alphabet is $Y = \{0, 1\}$, having the following channel matrix where ϵ is the probability of transmission error:

$$\begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}$$

(*i*) How much uncertainty is there about the input symbol once an output symbol has been received? [5 marks]

(*ii*) What is the mutual information $I(X;Y)$ of this channel? [2 marks]

(*iii*) What value of ϵ maximises the uncertainty $H(X|Y)$ about the input symbol given an output symbol? [1 mark]

(*b*) For a continuous (i.e. non-discrete) function $g(x)$, define:

(*i*) its continuous Fourier transform $G(k)$; [2 marks]

(*ii*) the inverse Fourier transform that recovers $g(x)$ from $G(k)$. [2 marks]

(*c*) What simplifications occur in the Fourier representation of a function if:

(*i*) the function is real-valued rather than complex-valued? [1 mark]

(*ii*) the function has even symmetry? [1 mark]

(*iii*) the function has odd symmetry? [1 mark]

(*d*) Give a bit-string representation of the number 13 in

(*i*) unary code for non-negative integers; [1 mark]

(*ii*) Golomb code for non-negative integers with parameter $b = 3$; [2 marks]

(*iii*) Elias gamma code for positive integers. [2 marks]

9 Types

(*a*) State the *value-restricted* typing rule for let-expressions in ML. [5 marks]

(*b*) Which of the following typing judgements are provable in the ML type system with the value-restricted rule for let-expressions? Justify your answer in each case, stating any other of the ML typing rules that you use. For part (*iii*) you must decide whether or not there is a type scheme σ that makes the typing provable.

(*i*) $\{\,\} \vdash \texttt{let}\ r = \texttt{ref}\ \lambda x(x)\ \texttt{in}\ (!r)(r := \lambda y(\texttt{true})) : \mathit{unit}$ [5 marks]

(*ii*) $\{\,\} \vdash \texttt{let}\ r = \texttt{ref}\ \lambda x(x)\ \texttt{in}\ (!r)(r := \lambda y(())) : \mathit{unit}$ [3 marks]

(*iii*) $\{\,\} \vdash \texttt{let}\ f = \lambda x(\texttt{ref}\ x)\ \texttt{in}\ f\,f : \sigma$ [3 marks]

(*iv*) $\{x : \alpha\} \vdash \lambda f(f\,x) : \forall\beta((\alpha \to \beta) \to \beta)$ (where α and β are distinct type variables) [2 marks]

(*v*) $\{x : \beta\} \vdash \lambda f(f\,x) : \forall\beta((\beta \to \beta) \to \beta)$ [2 marks]

10 Digital Signal Processing

(*a*) Characterise the systems below as linear/non-linear, causal/non-causal, and time invariant/time varying:

(*i*) $y_n = ax_{3n-2}$ [2 marks]

(*ii*) $y_n = y_{n-1} + 6x_{n-2}$ [2 marks]

(*iii*) $y_n = y_{n-1} - x_{n+5} + x_{n-5}$ [2 marks]

(*iv*) $y_n = \frac{x_n}{x_{n-3}y_{n-2}}$ [2 marks]

(*v*) $y_n = x_n - \cos\left(\frac{\pi}{2}n\right)$ [2 marks]

(*b*) Consider the system $h : \{x_n\} \to \{y_n\}$ with $y_n - y_{n-1} = x_n - x_{n-3}$.

(*i*) Give the impulse response of this system. [2 marks]

(*ii*) Give one sine-wave input sequence of the form

$$x_n = a \cdot \sin(b \cdot n + c)$$

(with $a \neq 0, b \neq 0$) for which $y_n = 0$ for all n. [2 marks]

(*iii*) Express the system h as a rational function $H(z)$. [3 marks]

(*iv*) Determine the values $z \in \mathbb{C}$ for which $H(z) = 0$. [3 marks]

11 Topics in Concurrency

(*a*) You are reminded that the logic CTL has assertions of the form

$$\mathbf{EX}\ A\ ,\ \mathbf{EG}\ A\ ,\ \mathbf{E}[A_0\ \mathbf{U}\ A_1]\ .$$

Explain their semantics in terms of paths. Describe their translation into the modal μ-calculus with a single action label. [6 marks]

(*b*) Give a finite assertion A in Hennessy–Milner logic with the following property

$$p \models A \ \text{ iff } \ p \text{ is strongly bisimilar to the CCS process } a.\mathbf{nil}\ ,$$

for any CCS process p with actions restricted to being within the set $\{a, b\}$. [7 marks]

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

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

Define $t \leqslant u$ iff there is a simulation S with $(t, u) \in S$.
Exhibit two CCS terms t and u for which $t \leqslant u$ and $u \leqslant t$ and yet where t and u are not strongly bisimilar. Briefly justify your answer. [7 marks]

12 Information Retrieval

The SNOWBALL algorithm uses bootstrapping from known tuples of named entities which stand in a well-defined relationship, in order to detect new tuples.

(*a*) Describe SNOWBALL's algorithm in detail, including the thresholds used in the single steps of the algorithm. [7 marks]

(*b*) The table below contains corpus examples of co-occurrences of organisation names (o) and location names (l). Consider a situation where SNOWBALL is applied to the corpus examples given here, when the only known tuples are <Microsoft, Redmond> and <Exxon, Irving>.

A	<l>Seattle</l>-based company <o>Boeing</o> offered ...
B	Yesterday, at <o>Microsoft</o>'s headquarters in <l>Redmond</l>, the deal was brokered ...
C	Though they had never been at <l>Redmond</l>, <o>Microsoft</o> showed them ...
D	In <l>New York</l>, <o>Microsoft</o> stock nosedived ...
E	When we arrived in <l>London</l>, <o>Exxon</o>petrol stations were ...
F	... met at <o>Microsoft</o> headquarters. In <l>Redmond</l>, ...
G	<o>Boeing</o>, <l>Seattle</l>, had no choice but to ...
H	In <l>New York</l>, <o>Intel</o> stock recovered ...
I	... due to arrive in <l>Irving</l>, <o>Exxon</o> executives might ...
J	<o>Boeing</o> headquarters in <l>Seattle</l> are air-conditioned ...
K	<o>Microsoft</o>, <l>Redmond</l>, made a statement ...
L	<o>Boeing</o>, <l>Seattle</l>, confirmed ...
M	<o>Microsoft<o>, <l>Redmond</l>, readily agreed ...
N	... <o>Exxon</o>. Although they had never in their whole life been in <l>Irving</l>, they ...
O	<o>Exxon</o>, <l>New York</l>, was a winner in our recent ...

Discuss which patterns get hypothesised and which new tuples this produces in the next iteration. Assume sensible thresholds. [6 marks]

(*c*) What happens to the result in part (*b*) if the sentence "*Microsoft's previous headquarters in Cincinnati were insured for 20 million dollars.*" gets added to the corpus? [3 marks]

(*d*) The SNOWBALL algorithm is to be applied to find tuples of person names and their professional positions from a large newspaper corpus. Would you expect SNOWBALL to work well on this task, and why? [4 marks]

13 Business Studies

(*a*) Explain the difference between *debt* and *equity* financing. [5 marks]

(*b*) What is meant by a *convertible debenture with coupon*, and how does this differ from a *redeemable preference share*? [5 marks]

(*c*) A certain software company has assets valued at £1M and is making a net profit of £100K on a turnover of £1M growing at 10% per annum. 1m shares, including 100K staff options, have been issued. No dividends have been paid. The company wish to raise money so that they can expand faster, and are prepared to sell up to 300K new shares. Price the issue. [5 marks]

(*d*) Why do venture capital funds prefer to invest in preference shares? [5 marks]

14 E-Commerce

(*a*) Consider setting up a rival to iTunes, the popular site where music tracks can be searched, purchased and legally downloaded. What business models might be appropriate? [5 marks]

(*b*) Outline the system design you would advise, with rough estimates of scale. [5 marks]

(*c*) What major difficulties will the project have to overcome? [5 marks]

(*d*) Can this model be applied to other areas, such as for example recipes from published cookbooks? Provide a SWOT analysis of the opportunity. [5 marks]

15 Additional Topics

(*a*) Describe briefly the concepts of throughput and latency in networking design. Does high latency always imply low throughput or *vice versa*? Explain your answer. [4 marks]

(*b*) Detail how a traditional operating system processes TCP/IP packets from a Network Interface Card (NIC). Discuss recent efforts to improve the latency of traditional networking operations. [8 marks]

(*c*) Explain how data is transferred from one machine to the next using the Shared Memory Model (SMM). [4 marks]

(*d*) Discuss the problem of application synchronisation using the SMM, highlighting different solutions that have been attempted to overcome this problem. [4 marks]

16 Additional Topics

(*a*) The Global Positioning System (GPS) is often described in terms of three Segments: Space, Control and User. Give brief descriptions of the components involved in each of these segments, and how they work together. [5 marks]

(*b*) Explain what is meant by the term *pseudo-range*. If a single pseudo-range is measured from a pair of satellites well above the horizon, the possible receiver positions will all lie close to an imaginary surface. Illustrate with a simple (2-D) diagram what a slice through this surface in the plane of the receiver and the satellites might look like. Show how errors in the pseudo-range will expand the possible receiver positions. Explain why this geometry might lead to large errors if a position is calculated from four satellites well above the horizon. [6 marks]

(*c*) Differential GPS is a technique for removing some of the errors inherent in the GPS. Explain how the errors might arise, how they are measured at a single base station, and how they can be distributed to users of the system. [5 marks]

(*d*) After Differential GPS has been used to remove errors, explain what error sources might be left, and how, if at all, they could be reduced. [4 marks]

END OF PAPER