COMPUTER SCIENCE TRIPOS Part II

Thursday 6 June 2002 1.30 to 4.30

Paper 9

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 Denotational Semantics

(*a*) Let A be a set. Let R consist of a set of pairs (X_0, a) where X_0 is a finite subset of A and $a \in A$. Define a function on the powerset of A,

$$\widehat{R} : \mathcal{P}(A) \to \mathcal{P}(A)$$

by

$$\widehat{R}(X) = \{\ a \mid\ \exists X_0 \subseteq X.\ (X_0, a) \in R\}$$

for $X \subseteq A$. Prove that the function $\widehat{R}$ is a continuous function on the domain $(\mathcal{P}(A), \subseteq)$. [6 marks]

(*b*) Describe how to construct the function cpo $((D \to E), \sqsubseteq)$ of two cpos $(D, \sqsubseteq_D)$ and $(E, \sqsubseteq_E)$. Prove that $((D \to E), \sqsubseteq)$ is a cpo.
(You may use facts about least upper bounds provided you state them clearly.) [7 marks]

(*c*) Exhibit two terms of PCF which are contextually equivalent and yet have distinct denotations in the domain $(\mathbb{B}_\perp \to (\mathbb{B}_\perp \to \mathbb{B}_\perp)) \to \mathbb{B}_\perp$ where $\mathbb{B} = \{\mathit{true}, \mathit{false}\}$ is the set of truth values. Exhibit the domain element on which the two denotations differ. [7 marks]

2 VLSI Design

Write brief notes (including circuit diagrams if appropriate) on the following schemes for carry propagation in binary addition:

(*a*) ripple carry;

(*b*) Manchester carry;

(*c*) carry look-ahead tree;

(*d*) carry skip;

(*e*) carry select.

[4 marks each]

3 Digital Communication II

(*a*) Current TCP incurs loss to discover the available capacity. Describe the AIMD (Additive Increase, Multiplicative Decrease) mechanism, and fast recovery, and show how this leads to the characteristic "saw tooth" throughput behaviour of TCP over time. [5 marks]

(*b*) Consider a TCP connection operating in steady-state whereby each time the congestion window increases to W segments a single packet loss occurs. In terms of W and the round trip time (R), derive a simple formula for the time between the minimum and maximum data rates achieved. In this optimal scenario, how many packets are sent between each loss event? [4 marks]

(*c*) Derive the connection's average throughput in terms of the fraction of packets lost (p), the connection's round trip time (R), and the segment size (B). [4 marks]

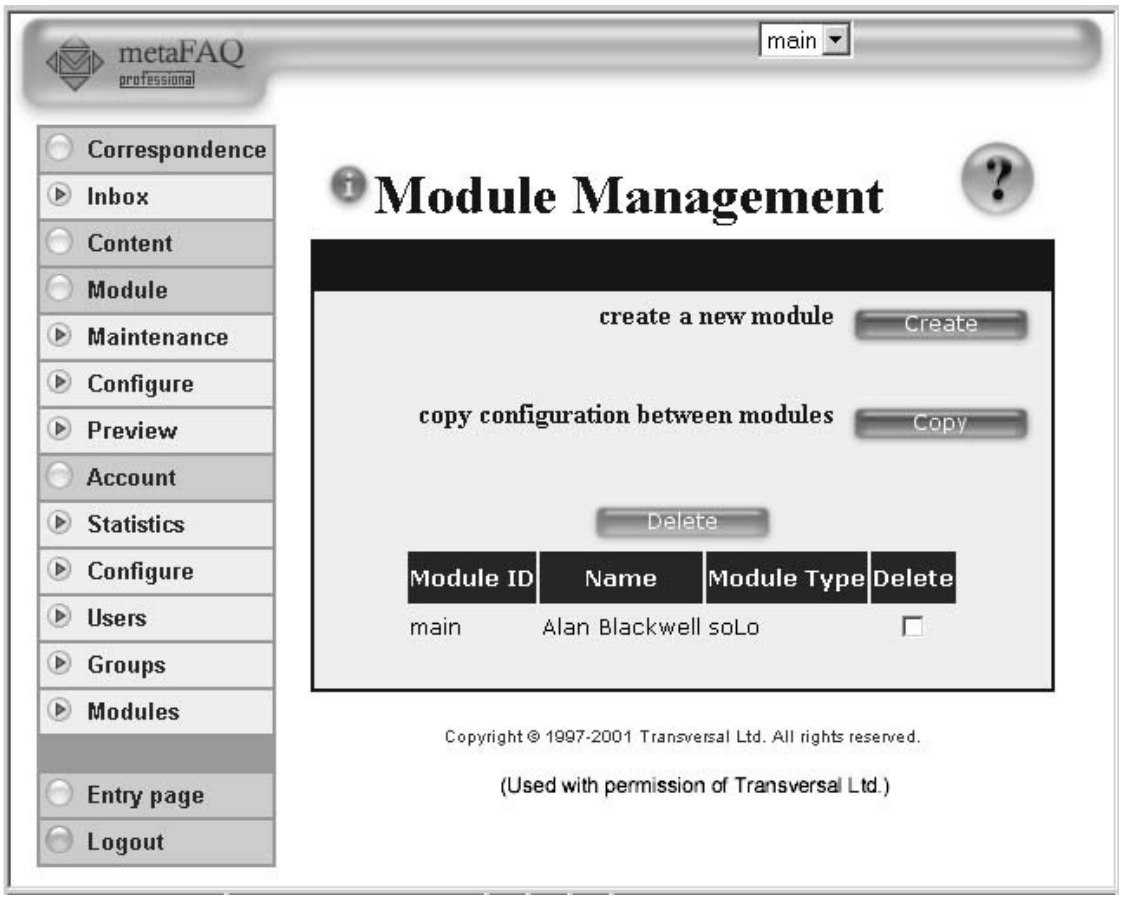
(*d*) Under what conditions is this very simplistic model likely to be accurate? [3 marks]

(*e*) How can router support for Explicit Congestion Notification (ECN) be used to smooth the TCP throughput? [4 marks]

4 HCI

(*a*) Identify *three* notations that might be found in the user interface of a generic word processor. For *each* of these, identify an important attribute relevant to the design of a new word processor, explaining how *each* would affect users' experience of the product. [10 marks]

(*b*) The following screen shot is taken from a product designed by a local company.



Describe *two* techniques for examining usability of this product, suggesting for *each* of them *four* ways in which they might influence the design. [10 marks]

5 Business Studies

(*a*) Explain the differences between

(*i*) credit and debit;

(*ii*) cash-flow and profit & loss statements;

(*iii*) equity and debt finance;

(*iv*) NPV and IRR;

(*v*) asset and DCF based valuation.

[2 marks each]

(*b*) A certain small software company has assets of about £100K (not including development work-in-progress), and an average cash-flow of about £15,000 per month, with a net profit of around £2000 per month. They are developing, but have not yet completed, a new graphical search engine into which they have invested about £100K of design and programmer time. The founders have invested about £150K, mostly in equity, and there is a long term debenture of £100K.

Provide a range of valuations for the company. Include notes explaining your assumptions and the basis for each valuation. [10 marks]

6 Types

(a) Give the typing rules for the polymorphic lambda calculus (PLC). [5 marks]

(b) Let $prod(\alpha_1, \alpha_2)$ denote the PLC type $\forall\alpha((\alpha_1 \to (\alpha_2 \to \alpha)) \to \alpha)$. Explain how it behaves like the ML product type $\alpha_1 * \alpha_2$. To do so, you should give PLC expressions *pair*, *fst* and *snd* of types

$$\begin{aligned} & \forall\alpha_1(\forall\alpha_2(\alpha_1 \to (\alpha_2 \to prod(\alpha_1, \alpha_2)))), \\ & \forall\alpha_1(\forall\alpha_2(prod(\alpha_1, \alpha_2) \to \alpha_1)), \\ \text{and} \quad & \forall\alpha_1(\forall\alpha_2(prod(\alpha_1, \alpha_2) \to \alpha_2)) \end{aligned}$$

respectively, corresponding to the ML polymorphic pairing and projection functions

```
fn x1 => fn x2 => (x1, x2) ,
fn (x1, x2) => x1 ,
and fn (x1, x2) => x2 .
```

Give proofs for the typing of *pair* and *fst*, and explain the beta-conversion properties of $fst\,\tau_1\,\tau_2(pair\,\tau_1\,\tau_2\,M_1\,M_2)$, for any PLC types τ_1, τ_2 and terms M_1, M_2. [10 marks]

(c) Is it always the case that a PLC term M of type $prod(\tau_1, \tau_2)$ is beta-convertible to $pair\,\tau_1\,\tau_2(fst\,\tau_1\,\tau_2\,M)(snd\,\tau_1\,\tau_2\,M)$? Justify your answer. [Hint: consider terms with free variables.] [5 marks]

7 Optimising Compilers

(*a*) Explain clearly an algorithm to *schedule* assembler code for a given architecture to increase its efficiency; the algorithm must give exactly the same instructions as in the original code, possibly in a different order. You may assume the hardware is interlocked so that NOPs are not required. Your answer should state the range over which scheduling takes place and, if this is smaller than a procedure, what actions can sensibly be taken at the boundary between separately scheduled sections of code. [10 marks]

(*b*) Consider an architecture which is ARM-like—the first operand is the destination register except in the case of stores. All operations take one cycle, with the exceptions that (*i*) using the value resulting from an immediately previous load will cause a single-cycle stall, as does (*ii*) the second of two successive store instructions. Use your algorithm to schedule the following code, noting briefly, for each instruction emitted, why your algorithm has chosen this instruction over other candidates.

```
ldr  r3,[r1,#0]
str  r3,[r2,#0]
ldr  r4,[r1,#4]
str  r4,[r2,#4]
add  r5,r4,#4
```

[6 marks]

(*c*) Give with brief explanation the number of possible valid schedulings of the following code (not just the one produced by your algorithm):

```
add  r3,r1,r2
or   r4,r1,r2
sub  r5,r3,r4
xor  r4,r1,r3
```

[4 marks]

8 Advanced Algorithms

(*a*) Explain how to check a large number for primality using a probabilistic method that gives you a bound of the probability of getting an incorrect judgment. [7 marks]

(*b*) Give an asymptotic formula predicting the number of computer operations needed to verify that a number with n bits is prime, supposing that multiplication, division and remaindering are done using $O(n^2)$ methods and that you want to achieve a probability of error bounded by 1 in 2^{60}. You do not need to prove that the algorithm you describe works, but you should nevertheless explain it carefully and completely. [7 marks]

(*c*) The gap between adjacent primes near the integer N is roughly $\log(N)$. Estimate roughly the number of computer operations you would expect to be needed to find a 2000-bit prime that is just slightly larger than some given 2000-bit random number. [6 marks]

9 Neural Computing

Explain the mechanisms and computational significance of nerve impulse generation and transmission. Include the following aspects:

(a) Equivalent electrical circuit for an excitable nerve cell membrane.

(b) How different ion species flow across the membrane, in terms of currents, capacitance, conductances, and voltage-dependence. (Your answer can be qualitative.)

(c) Role of positive feedback and voltage-dependent conductances.

(d) The respect in which a nerve impulse is a mathematical catastrophe.

(e) Approximate time-scale of events, and the speed of nerve impulse propagation.

(f) What happens when a propagating nerve impulse reaches an axonal branch.

(g) What would happen if two impulses approached each other from opposite directions along a single nerve fibre and collided.

(h) How linear operations like integration in space and time can be combined in dendritic trees with logical or Boolean operations such as AND, OR, NOT, and veto.

(i) Whether "processing" can be distinguished from "communications" as it is for artificial computing devices.

(j) Respects in which stochasticity in nerve impulse time-series may offer computational opportunities that are absent in synchronous deterministic logic.

[2 marks each]

10 Information Theory and Coding

(*a*) (*i*) A Hamming Code allows reliable transmission of data over a noisy channel with guaranteed error correction as long as no more than one bit in any block of 7 is corrupted. What is the maximum possible rate of information transmission, in units of (data bits reliably received) per (number of bits transmitted), when using such an error correcting code? [2 marks]

(*ii*) In such a code, what type of Boolean operator on the data bits is used to build the syndromes? Is this operator applied before transmission, or upon reception? [2 marks]

(*b*) (*i*) For each of the four classes of signals in the following table,

Class	*Signal Type*
1.	continuous, aperiodic
2.	continuous, periodic
3.	discrete, aperiodic
4.	discrete, periodic

identify its characteristic spectrum from the following table:

Class	*Spectral Characteristic*
A.	continuous, aperiodic
B.	continuous, periodic
C.	discrete, aperiodic
D.	discrete, periodic

(Give your answer just in the form 1-A, 2-B, etc. Note that you have 24 different possibilities.) [8 marks]

(*ii*) For each case, name one example of such a function and its Fourier transform. [4 marks]

(*c*) Give two reasons why Logan's Theorem about the richness of zero-crossings for encoding and recovering all the information in a one-octave signal may not be applicable to images as it is for one-dimensional signals. [4 marks]

11 Numerical Analysis II

Consider the alternative formulae

$$y_{n+1} = y_n + hf(x_n, y_n) + O(h^2) \tag{1}$$

$$y_{n+1} = y_{n-1} + 2hf(x_n, y_n) + O(h^3) \tag{2}$$

applied to the ODE

$$y' = -5y, \quad y(0) = 1$$

using $h = 0.1$ in each case.

(a) Define the terms *local error* and *order* for an ODE formula. What is the *order* of each of the methods (1) and (2)? [2 marks]

(b) Giving answers to 2 significant decimal digits of accuracy, compute the solution of the ODE for $x_n = 0, 0.1, 0.2, \ldots\ 1.0$ for each method. Tabulate your answers. The exact solutions to 2 significant digits are:

$$1.0,\ 0.61,\ 0.37,\ 0.22,\ 0.14,\ 0.082,\ 0.050,\ 0.030,\ 0.018,\ 0.011,\ 0.0067$$

Assume the exact value of $y(0.1)$ for method (2). [7 marks]

(c) Which method is more accurate initially and why? Explain the behaviour of each method as x increases. [3 marks]

(d) Solve the ODE. Find a general term for y_n in method (1) and show that the absolute error in (1) will be small when n is large. Without performing any further calculations, how do you expect the absolute error in method (2) to behave when n is large? [5 marks]

(e) Discuss briefly the suitability of formulae (1) and (2) as predictors for predictor–corrector methods in respect of *order* and *stability*. [3 marks]

12 Specification and Verification II

The multiplexer `MUX`, register `REG c` (where `c` is the initial value) and combinational unit `COM f` (where `f` is the function computed) are defined to have the behaviour given below.

```
MUX(sw,i1,i2,o) = ∀t. o t = if sw t then i1 t else i2 t
REG c (i,o)     = (o 0 = c) ∧ ∀t. o(t+1) = i t
COM f (i,o)     = ∀t. o t = f(i t)
```

Using only instances of `MUX`, `REG c` and `COM f` design a device `DEV(c,f)` that satisfies

```
DEV(c,f)(reset,i,o) =
  (o 0 = c) ∧  ∀t. o(t+1) = if reset(t+1) then c else f(o t)
```

[8 marks]

Prove that your design meets this specification. [12 marks]

13 Computer Vision

(*a*) Consider the "eigenfaces" approach to face recognition in computer vision.

(*i*) What is the rôle of the database population of example faces upon which this algorithm depends? [4 marks]

(*ii*) What are the features that the algorithm extracts, and how does it compute them? How is any given face represented in terms of the existing population of faces? [4 marks]

(*iii*) What are the strengths and the weaknesses of this type of representation for human faces? What invariances, if any, does this algorithm capture over the factors of perspective angle (or pose), illumination geometry, and facial expression? [4 marks]

(*iv*) Describe the relative computational complexity of this algorithm, its ability to learn over time, and its typical performance in face recognition trials. [4 marks]

(*b*) What is the following block of code doing over the image array `image[i][j]` as it computes the resulting new image array `result[i][j]` ? Give the appropriate mathematical name for this operation, and describe what it accomplishes. What are some computer vision tasks that might use this block of four nested `for loops`?

```
for (i = 0; i < iend; i++) {
    for (j = 0; j < jend; j++) {
        sum = 0;
        for (m = 0; m < mend; m++) {
            for (n = 0; n < nend; n++ ) {
                sum += kernel[m][n] * image[i-m][j-n];
            }
        }
        result[i][j] = sum/(mend*nend);
    }
}
```

[4 marks]

14 Natural Language Processing

The Ultimate Dating Agency has a database which contains enough information to respond to requests like:

> List every computer scientist with any friends who obsess(es) about some reprogrammable device(s).

(*a*) Describe the different queries that result from adding each morphological suffix in brackets. [2 marks]

(*b*) Give formulae of first-order logic which represent these different queries. [6 marks]

(*c*) Describe techniques for morphological analysis, syntactic parsing and compositional semantic interpretation that would output such representations. [9 marks]

(*d*) How might the database information be represented and queried? [3 marks]

15 Topics in Concurrency

This question assumes familiarity with the process language **SPL** and its event-based semantics. Suppose that there are three agents A, B, S which run at most once (so with no replication) according to the scheme:

$$
\begin{aligned}
&A \to B:\ A \\
&B \to A:\ n \quad \text{(a nonce)} \\
&A \to B:\ \{n\}_{Key(A,S)} \\
&B \to S:\ \{A,\ \{n\}_{Key(A,S)}\}_{Key(B,S)} \\
&S \to B:\ \{A,\ n\}_{Key(B,S)}
\end{aligned}
$$

The symmetric key $Key(A, S)$ is shared between A and S, and the symmetric key $Key(B, S)$ between B and S.

(a) Express the agents A, B and S as processes in **SPL**. [3 marks]

(b) Describe diagrammatically the reachable events of A, B and S, taking care to specify the pre- and postconditions, and actions of the events (you may, however, omit descriptions of the control conditions). [5 marks]

(c) The agents communicate in the presence of an attacker which can decrypt and encrypt messages with available keys, and compose and decompose messages. Draw the four kinds of events (decryption, encryption, composition, decomposition), taking care to specify the nature of their pre- and postconditions. [2 marks]

Let P be the Petri net obtained as the union of the events of A, B, S and the attacker.

(d) Let $Q(\mathcal{M})$ be the property of net P's markings $\mathcal{M}$ which holds when no output message in $\mathcal{M}$ has either key $Key(A, S)$ or $Key(B, S)$ as a submessage. Explain why the property Q holds at all reachable markings of P provided it holds at its initial marking. [6 marks]

(e) Assume that Q holds at the initial marking. Draw, without proof, the dependency on events of agents A, B and S which shows that if B inputs a message $\{A,\ n\}_{Key(B,S)}$, then A has previously input a message n. [4 marks]

END OF PAPER