COMPUTER SCIENCE TRIPOS Part II

Tuesday 4 June 2002 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 Specification and Verification I

(*a*) Describe the difference between deep and shallow semantic embedding. [4 marks]

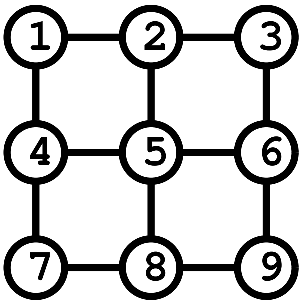
(*b*) Describe *two advantages* of using deep embedding and *two disadvantages*. [4 + 4 marks]

(*c*) Outline how partial and total correctness specifications can be translated into higher order logic. [4 marks]

(*d*) Give *one advantage* and *one disadvantage* of regarding Hoare Logic as a theory in higher order logic. [2 + 2 marks]

2 Specification and Verification II

Consider a 3 × 3 array of 9 switches



Suppose each switch $1, 2, \ldots, 9$ can be either on or off, and that toggling any switch will automatically toggle all its immediate neighbours. For example, toggling switch 5 will also toggle switches 2, 4, 6 and 8, and toggling switch 6 will also toggle switches 3, 5 and 9.

(*a*) Devise (*i*) a state space and (*ii*) transition relation to represent the behaviour of the array of switches. [4 + 6 marks]

(*b*) You are given the problem of getting from an initial state in which even numbered switches are on and odd numbered switches are off, to a final state in which all the switches are off.

Write down predicates on your state space that characterise the (*i*) initial and (*ii*) final states. [2 + 2 marks]

(*c*) Explain how you might use a model checker to find a sequence of switches to toggle to get from the initial to final state. [6 marks]

You are not expected actually to solve the problem, but only to explain how to represent it in terms of model checking.

3 Comparative Architectures

A naïve programmer writes the following code for performing the matrix multiply-add function $C = AB + C$ on square matrices:

```
for (i=0;i<N;++i) {
  for(j=0;j<N;++j) {
    for(k=0;k<N;++k) {
       C[k][i] = C[k][i] + ( A[k][j] * B[j][i] );
    }
  }
}
```

(where X[v][u] refers to the element in row v, column u. Arrays are stored in memory row by row, i.e.
X[0][0],X[0][1],X[0][2]...X[0][N],X[1][0],X[1][1], etc.)

(*a*) When used to multiply very large matrices, performance of the programmer's algorithm is very poor. Explain what is happening. [6 marks]

(*b*) The algorithm can be improved simply by changing the order of the loops. Demonstrate how and why. [5 marks]

(*c*) Show how further improvement can be obtained through a technique known as *cache blocking*. [5 marks]

(*d*) Could the algorithm be successfully parallelised to run on a microprocessor supporting Simultaneous Multithreading (SMT)? Briefly justify your answer. [4 marks]

4 Optimising Compilers

(*a*) Explain how the *register interference graph* (sometimes known as the *clash graph*) is constructed and briefly explain its relevance to register allocation. [5 marks]

(*b*) Determine the register interference graph for the following flowgraph program expressed in an assembly language notation. Only lines 3 and 10 have labels; the remaining lines are numbered for your convenience. The compare and branch instructions are written on a single line to reflect the flowgraph view of them. Give the interference graph in diagrammatic form, and annotate each edge in it with the line or lines in the program which produce it.

```
        Length:
L0              MOV   p, r0
L1              MOV   i, #0
L2              CMP   p, #0;  BEQ   L10
        L3:     MOV   t3, i
L4              ADD   t4, t3, #1
L5              MOV   i, t4
L6              MOV   t5, p
L7              LDR   t6, [t5, #4]
L8              MOV   p, t6
L9              CMP   t6, #0;  BNE   L3
        L10:    MOV   r0, i
L11             RET
```

[Hint: you may wish to construct a table showing which locations have which variables live at that point.] [10 marks]

(*c*) Suppose a given flowgraph program uses k registers when a given compiler colours its variables with registers. Now suppose the program is converted to *static single assignment* (SSA) form and the latter coloured by the same compiler, this time using l registers. How would you expect k and l to compare? Discuss, giving reasons, whether there might be some program which disagreed with your expectation. [5 marks]

5 E-Commerce

A new start-up company proposes to develop an electronic wallet, a device that can cryptographically hold electronic money, data, credit card numbers etc. Such a device might, for example, be included in a mobile phone.

(*a*) Explain how network externalities affect the introduction of such a device. [5 marks]

(*b*) Explain some of the legal and regulatory issues affecting such a device. [5 marks]

(*c*) Sketch out the back-end processing and infrastructure that would be needed to support such a device. [5 marks]

(*d*) Would such a device increase the overall security of a transaction? Justify your answer. [5 marks]

6 Security

(*a*) One of the algorithms that you implement inside a smartcard requires a stack (LIFO) for storing data records. Assume that each record is a few hundred bytes long. This stack can become very large and the small amount of memory available in the card is not sufficient to hold it. Therefore you decide that the stack object will be implemented via remote method invocation in the card terminal, which has enough memory.

The externally stored stack offers the usual four methods: `init` to initialise an empty stack, `push` to place a data record onto the stack, `isempty` to test whether the stack contains no record, and `pop` to retrieve the top record from the stack.

The integrity of the stack is crucial for the security of the application and the card terminal is not tamper resistant. Consider an algorithm for an integrity-protected stack object that will be implemented on the smartcard and uses an existing secure hash function h available in the card as well as the external stack object.

(*i*) Why is just adding a simple message authentication code to each externally stored record not sufficient here? [2 marks]

(*ii*) What check data do you attach to externally stacked records, such that the memory required in the smartcard does not grow with the stack size? What check data remains inside the card? Show the resulting internal check values and the records on the external stack (call them $Y_1, Y_2, \ldots$) after you pushed three data records X_1, X_2, X_3 onto the stack. [6 marks]

(*iii*) Write short pseudo-code for the methods of the protected stack object (`secure_init`, `secure_push`, `secure_pop`, `secure_isempty`) that shows how they update the on-card check data and under which conditions a tampering alarm is raised. [6 marks]

(*b*) In the same smartcard application, you also need an externally stored integrity-protected queue (FIFO). You decide to protect each externally stored record with a MAC for which a new key will be generated whenever the FIFO is initialised. What check data beyond the MAC key needs to be kept inside the card? What additional check data do you have to add to the records to guarantee the integrity of the FIFO? [6 marks]

7 Neural Computing

(*a*) When using a feed-forward neural network to solve a classification problem, one alternative is to interpret the network's outputs as posterior probabilities of class membership, and then to use these posterior probabilities to make classification decisions. Alternatively, we can treat the network as a discriminant function which is used to make the classification decisions directly. Discuss the relative merits of these two approaches. [8 marks]

(*b*) Explain the concept of a likelihood function, and the principle of maximum likelihood in model building and inference from data. [2 marks]

(*c*) Explain the key ideas of a Hopfield neural network for content-addressable, associative memory. In explaining how memories are stored and retrieved, be sure to define the notions of:

- configuration space
- connectivity matrix
- stable attractor
- basin of attraction
- network capacity, and its dependence on the number of "neurones"

[10 marks]

8 Computer Systems Modelling

Consider an $M/M/1$ queue and represent the state of the queue by the number of customers present.

(a) Draw a state diagram for the Markov chain describing the state of the queue showing the possible states and transition rates. Briefly explain the diagram. [4 marks]

(b) What is the condition for the existence of a steady-state equilibrium distribution p_k $(k = 0, 1, 2, \ldots)$ for the number present? [2 marks]

(c) Determine the steady-state average number of customers present. [4 marks]

(d) Use Little's law to determine the steady-state average response time that a customer spends in the system. [2 marks]

(e) Suppose that a single communication channel is used to carry data items sent by various sources connected to the channel. Assume that each source generates a stream of data items with inter-arrival times which are exponential at rate 2 items/second and that all sources are statistically independent. All the items wait in a single queue and are transmitted one at a time. The transmission times are exponentially distributed with mean 25 milliseconds and are statistically independent. Determine the largest number of sources that can be connected to the channel according to each of the following two criteria:

(i) The channel is not saturated.

(ii) The steady-state average response time for an item must not exceed 100 milliseconds.

[8 marks]

9 Advanced Graphics

(*a*) A disc is a finite, planar, circular object. Describe an algorithm to find the point of intersection of an arbitrary ray with an arbitrary disc in three dimensions. Ensure that you describe the parameters used to define both the ray and the disc. [6 marks]

(*b*) Given the above algorithm and an algorithm to find the intersection of an arbitrary ray with a finite-length *open* cylinder, a programmer has two choices for implementing an algorithm to find the intersection with a finite-length *closed* cylinder. She could simply use the finite-length open cylinder primitive and two disc primitives. Alternatively she could implement the finite-length closed cylinder as a primitive in its own right by adding extra code to the open cylinder algorithm. Compare the two alternatives in terms of efficiency and accuracy. [4 marks]

(*c*) Describe the situations in which it is sensible to use a winged-edged data structure to represent a polygon mesh and, conversely, the situations in which a winged-edged data structure is not a sensible option for representing a polygon mesh. What is the minimum information which is required to successfully draw a polygon mesh using Gouraud shading? [4 marks]

(*d*) Derive the formula of and sketch a graph of $N_{3,3}(t)$, the third of the quadratic B-spline basis functions, for the knot vector $[\,0\,0\,0\,1\,3\,3\,4\,5\,5\,5\,]$. [6 marks]

10 HCI

(*a*) Describe *two* quantitative and *two* qualitative techniques for analysing the usability of a software product. [4 marks]

(*b*) Compare the costs and benefits of the quantitative techniques. [6 marks]

(*c*) Compare the costs and benefits of the qualitative techniques. [6 marks]

(*d*) If restricted to a single one of these techniques when designing a new online banking system, which would you choose and why? [4 marks]

11 Natural Language Processing

An open-domain question-answering system should be able to answer questions such as the following by searching a database of documents and returning an appropriate text snippet. For example:

> User: What debts did Quintex leave?
>
> System 1: About $1.4 billion (Australian)

extracted from:

> Quintex Australia Ltd. and Quintex Ltd. together have debt of about $1.4 billion (Australian), according to two analysts at Australian brokerage firms.

and:

> System 2: Around ADollars 1.5bn (Pounds 680m)

from:

> Quintex group collapsed yesterday. The failure left corporate debts of around ADollars 1.5bn (Pounds 680m) and additional personal debts.

(*a*) What natural language processing techniques would be needed to build such a system? [8 marks]

(*b*) What problems would arise as a consequence of the open-domain requirement? [4 marks]

(*c*) What problems would arise selecting an appropriate text snippet from a matching sentence? [4 marks]

(*d*) How feasible would it be to have the system reliably rank such answers (in this case preferring the second)? [4 marks]

12 Information Theory and Coding

(*a*) State and explain (without proving) two different theorems about signal encoding that both illustrate the following principle: strict bandlimiting (either lowpass or bandpass) of a continuous signal reduces the information that it contains from potentially infinite to a finite discrete set of data, and allows exact reconstruction of the signal from just a sparse set of sample values. For both of your examples, explain what the sample data are, and why bandlimiting a signal has such a dramatic effect on the amount of information required to represent it completely. [10 marks]

(*b*) A variable length, uniquely decodable code which has the prefix property, and whose N binary code word lengths are

$$n_1 \leqslant n_2 \leqslant n_3 \cdots \leqslant n_N$$

must satisfy what condition on these code word lengths?

(State both the condition on the code word lengths, and the name for this condition, but do not attempt to prove it.) [4 marks]

(*c*) For a discrete data sequence consisting of the N uniformly-spaced samples

$$\{g_n\} = \{g_0,\ g_1,\ \ldots,\ g_{N-1}\}$$

define both the Discrete Fourier Transform $\{G_k\}$ of this sequence, and its Inverse Transform, which recovers $\{g_n\}$ from $\{G_k\}$. [6 marks]

13 Types

(*a*) State the typing rule for ML **let**-expressions, $\mathbf{let}\,x = M\,\mathbf{in}\,M'$, using typing judgements of the form $A, \Gamma \vdash M : \tau$ where A is a finite set of type variables, Γ is a finite function from variables to type schemes and τ is a type. [3 marks]

(*b*) Give an example to show that in a **let**-expression $\mathbf{let}\,x = M\,\mathbf{in}\,M'$, the **let**-bound variable x can occur polymorphically in the body M'. Give the proof of any valid typing judgement that you use. [5 marks]

(*c*) Give the ML typing rules for the unit value $\mathbf{()}$, for reference creation $\mathbf{ref}\,M$, for dereferencing $!M$, and for assignment $M := M'$. [4 marks]

(*d*) Explain how the combination of the typing rules from parts (*a*) and (*c*) leads to unsoundness of the type system. How does the revised definition of ML modify the typing rule for **let**-expressions in order to restore type soundness? [8 marks]

14 Additional Topics

You are required to design a Sentient Computing System using Active Bats to locate people and physical objects.

(*a*) Indicate how you would model the environment and represent space in your architecture. [8 marks]

(*b*) Show how you would use the tracked objects' dynamics to process and filter incoming location data. [8 marks]

(*c*) Discuss the feasibility of a spatial monitor that generates events in terms of geometric containment and overlapping. [4 marks]

15 Additional Topics

VNC is a protocol for remote access to graphical user interfaces.

(*a*) Describe the VNC architecture. [6 marks]

(*b*) Indicate what steps are taken to minimise network traffic. [5 marks]

(*c*) How does the protocol adapt to networks and clients of different speeds? [5 marks]

(*d*) Discuss how suitable VNC might be for used in third-generation mobile telephony (3G) systems. [4 marks]

END OF PAPER