COMPUTER SCIENCE TRIPOS Part II

Tuesday 7 June 2011 1.30 to 4.30

COMPUTER SCIENCE 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 cover sheets
Tags

SPECIAL REQUIREMENTS
Approved calculator permitted

1 Business Studies

(*a*) Distinguish between debt and equity finance. [5 marks]

(*b*) Under what conditions must the directors of a company declare insolvency? [5 marks]

(*c*) A small software company has the following simplified cashflow, funded by shareholders' equity of £20,000 and a bank overdraft of £5000:

Cashflow: Invoiced money received 2 months after due date; paid after 1 month of receipt of invoice

Month	1	2	3	4	5	6	7
Income			30,000				30,000
Expenditure							
Programmers	5,000	5,000	5,000	5,000	5,000	5,000	5,000
Overheads		5,000	5,000	5,000	5,000	5,000	5,000
Total costs	5,000	10,000	10,000	10,000	10,000	10,000	10,000
Cash at start	−5,000	15,000	5,000	25,000	15,000	5,000	−5,000
Cash at end	15,000	5,000	25,000	15,000	5,000	−5,000	15,000

Draw up the balance sheet at the end of month 6. [5 marks]

(*d*) What advice would you give the directors? [5 marks]

2 Artificial Intelligence II

Consider the following propositional planning problem.

Start state: $\neg A$, $\neg B$, $\neg C$, D.

Goal: A, B, C, $\neg D$.

Actions:

- Action 1 has preconditions A,B,C and effect $\neg D$.
- Action 2 has preconditions $\neg A$,$\neg B$ and effects A and B.
- Action 3 has preconditions $\neg B$,$\neg C$ and effects B and C.
- Action 4 has precondition B and effect $\neg B$.

(*a*) Using an entire sheet of paper, draw the planning graph as far as state level S_3, where the start state is at state level S_0 and the first action level is A_0. Do not add any mutex links at this point. [5 marks]

(*b*) Describe each of the five kinds of *mutex link* that can be incorporated in a planning graph. Add one example of each to the graph you produced in part (*a*). Clearly label the links to make clear which type they are. [10 marks]

(*c*) At which level in the planning graph will all goals first be present simultaneously? Will the GraphPlan algorithm be able to extract a working plan without extending it beyond this level? Explain your answer, adding further mutex links to your diagram if necessary. [5 marks]

3 Advanced Graphics

(*a*) Describe the key features of B-splines that make them useful for representing curves in computer-aided design (CAD). [3 marks]

(*b*) Derive and graph the quadratic Bézier basis functions using the standard B-spline method and the knot vector $[0, 0, 0, 1, 1, 1]$. [5 marks]

(*c*) Describe Chaikin's corner-cutting subdivision method that produces the quadratic Bézier curve in the limit. [5 marks]

(*d*) Describe the Doo–Sabin subdivision scheme, that is the generalisation of Chaikin's method to the bi-variate case able to represent surfaces with extraordinary vertices. [7 marks]

4 Comparative Architectures

(*a*) What hardware and software techniques may be used to reduce the number of conflict misses experienced by a direct-mapped cache? [4 marks]

(*b*) How might a hardware prefetcher that is capable of detecting and prefetching non-unit strides be implemented? [4 marks]

(*c*) How can the MESI cache coherency protocol be exploited to ensure that a test-and-set instruction is performed atomically without the need to lock down the bus for multiple cycles? [4 marks]

(*d*) Moore's law predicts we will be able to integrate a very large number of processor cores onto a single chip in the near future. What constraints and challenges may limit our ability to exploit these chip multiprocessors? [8 marks]

5 Denotational Semantics

(*a*) Let Ω be the domain

$$0 \sqsubseteq 1 \sqsubseteq \cdots \sqsubseteq n \sqsubseteq \cdots \sqsubseteq \omega \qquad (n \in \mathbb{N})$$

$\big($That is, $\Omega = \big(\mathbb{N} \cup \{\omega\}, \sqsubseteq\big)$ with $x \sqsubseteq y$ in Ω iff $x \leq y$ in $\mathbb{N}$ or $y = \omega$.$\big)$

Indicate whether the following statements are true or false. Provide an argument for each answer.

(*i*) Every monotone function from Ω to Ω is continuous. [5 marks]

(*ii*) Every monotone function from Ω to Ω has a least pre-fixed point. [5 marks]

(*b*) Let D and E be domains, and let $f : D \to D$ and $g : E \to E$ be continuous functions.

(*i*) Define $f \times g : D \times E \to D \times E$ to be the continuous function given by $\big(f \times g\big)(d, e) = \big(f(d), g(e)\big)$, and let $\pi_1 : D \times E \to D$ and $\pi_2 : D \times E \to E$ respectively denote the first and second projection functions.

Show that $\mathit{fix}(f \times g) \sqsubseteq \big(\mathit{fix}(f), \mathit{fix}(g)\big)$, and that $\mathit{fix}(f) \sqsubseteq \pi_1\big(\mathit{fix}(f \times g)\big)$ and $\mathit{fix}(g) \sqsubseteq \pi_2\big(\mathit{fix}(f \times g)\big)$. [5 marks]

(*ii*) It follows from part (*i*) that $\mathit{fix}(f \times g) = \big(\mathit{fix}(f), \mathit{fix}(g)\big)$. Use this and Scott's Fixed Point Induction Principle to show that, for all strict continuous functions $h : D \to E$,

$$h \circ f = g \circ h \implies h\big(\mathit{fix}(f)\big) = \mathit{fix}(g)$$ [5 marks]

6 Hoare Logic

(*a*) What does it mean for a logic to be *sound* and to be *complete*? [4 marks]

(*b*) Outline how Hoare logic can be proved sound. [4 marks]

(*c*) What does it mean for Hoare logic to be *relatively complete*? [4 marks]

(*d*) State conditions under which Hoare logic is relatively complete and outline the proof of completeness under these conditions. [8 marks]

7 Human–Computer Interaction

This question is concerned with the usability of software development tools. You are asked to compare tools based on UML diagrams with those based on the Java language. Consider in particular the definition of data structures using UML class diagrams, in contrast to equivalent class declarations in Java.

(*a*) Compare the purely graphical properties of these two notations, and the ways in which the graphical properties of each display correspond to the information structure being defined. Describe **three** ways in which the graphical correspondences are different. [2 marks each]

(*b*) What analytic approach could be used to compare the usability implications that arise from considerations such as those in part (*a*)? [1 mark]

(*c*) On the basis of the approach in part (*b*), suggest **three** specific usability issues relevant to the design of programming environments, in each case contrasting the implications for programming in UML and Java. [3 marks each]

(*d*) Describe an empirical approach that could be used to evaluate the usability of a new environment, with respect to one or more of the issues identified in part (*c*). Consider both the collection and analysis of data. [4 marks]

8 Mobile and Sensor Systems

Consider a vehicular network where cars are equipped with a Wi-Fi radio (802.11) interface. Assume a number of Wi-Fi basestations scattered around an area. The basestations provide only very sparse coverage of the area, but are connected to the Internet through a backbone network.

(*a*) Explain what effects the mobility of the vehicles can have on communication. In particular, describe short- and long-term fading effects. [3 marks]

(*b*) Illustrate the essential aspects of the MAC layer of 802.11. [3 marks]

(*c*) Consider RTS and CTS packets used in 802.11. Describe what problem they solve and indicate where this approach is incomplete. [2 marks]

(*d*) Assume that information about traffic needs to be disseminated from the basestations to all vehicles. Explain what routing protocol(s) would be suitable for a vehicular scenario like the one described (do not consider GPSR): justify your selection and the trade-offs of your choices. [6 marks]

(*e*) Now assume that vehicles collect traffic information (such as their speed and position) and want to forward this information to a server. Also assume that nodes adopt GPSR to route messages geographically to the closest basestation (position known). Explain how GPSR can be used for this purpose and under what conditions it would not work. [6 marks]

9 Natural Language Processing

(*a*) Give an equation for finding the most probable sequence of part of speech (POS) tags that could be utilised by a stochastic POS tagger. You should assume a bigram model. [5 marks]

(*b*) Given the following training data, show the estimates that would be obtained for the probabilities in the equation you gave:

```
the_DT0 green_AJ0 bottle_NN1 leaked_VVD ._PUN the_DT0
suppliers_NN2 bottle_VVB water_NN1 ._PUN green_AJ0 water_NN1
suppliers_NN2 bottle_VVB ._PUN
```

[4 marks]

(*c*) Explain what is meant by the terms *smoothing* and *backoff* in the context of stochastic POS tagging. [4 marks]

(*d*) One common source of errors in stochastic POS taggers is that nouns occurring immediately before other nouns (e.g. *catamaran trailer*) are often tagged as adjectives and, conversely, prenominal adjectives are often tagged as nouns (e.g. *trial offer*). Suggest possible reasons for this effect. [7 marks]

10 Temporal Logic and Model Checking

Write down temporal logic formulae that capture the meanings of the properties described in English below (**p** and **q** are atomic properties, i.e. properties of states). In each case say which logic you are using and briefly justify your choice.

(*a*) "on every path there is a point after which **p** is always true" [5 marks]

(*b*) "from every state it is possible to get to a state where **p** is true" [5 marks]

(*c*) "if **p** is true for 100 successive states then eventually **q** will be true" [5 marks]

(*d*) "if **p** is true infinitely often along a path then so is **q**" [5 marks]

11 Principles of Communications

(*a*) Consider a Graph G, with bi-directional equal weight edges, defined by:

$$\begin{aligned} G &= V, E \\ V &= 1, 2, 3, 4, 5, 6, 7 \\ E &= (2,1), (2,3), (2,4)(3,4)(5,2), (6,2), (7,1), (7,5), (7,6) \end{aligned}$$

(*i*) Draw the graph. [2 marks]

(*ii*) Define a spanning tree of G using an incidence matrix. [3 marks]

(*iii*) Explain the Distance Vector (D–V) routing algorithm, illustrating it operating in terms of messages sent and received by node 7 in G. Show what happens in terms of D–V messages exchanged by node 7, if the edge $(1, 7)$ breaks, and then later, when edge $(1, 7)$ is repaired. [10 marks]

(*b*) Suppose that your college has 1000 members; and that Cambridge has a population of 100,000; and that the Earth's total population is ten thousand million. Assuming that each of these social networks is a random graph with $p = 0.01$, then for each network, what is the average degree and average path length? [5 marks]

12 Security II

(*a*) Write out the Needham–Schroder protocol, explaining the notation you use. [5 marks]

(*b*) Describe the "bug" in the protocol, stating why some people consider it to be a bug and other people consider it not to be. [5 marks]

(*c*) Provide an amended protocol that does not have the "bug" but which (unlike Kerberos) uses random nonces rather than timestamps. [10 marks]

13 Optimising Compilers

(*a*) Define what is meant by a variable being *live* at a program point, carefully distinguishing versions based on program structure and I/O behaviour. Discuss these alternatives and their suitability for compilation using words like "decidable" and "safety". [4 marks]

(*b*) Explain the difference between *dead code* and *unreachable code*, mentioning any analysis necessary to remove such code. [3 marks]

(*c*) Let q be an expression (whose only free variable is `a`) which always evaluates to true although deducing this is beyond the power of a given compiler optimiser.

Consider the function

```
int p(int a, int b, int c, int d, int e)
{   int x = a+b, y = a+c, z=0, r;
    if (a<5) { r = a+x; while (1) continue; z = a+d; }
    else if (q) { r = y; }
    else { r = a+e; }
    return r+z;
}
```

(*i*) Give the flowgraph for `p` (assuming it to have been translated as naïvely as possible). [3 marks]

(*ii*) Using your definition of liveness suitable for a compiler from part (*a*), give the set of variables live on entry to `p`, comparing this with liveness required by I/O behaviour. [4 marks]

(*iii*) Expressing your answer in source code similar to the definition of `p`, give the code you might expect a good optimiser to achieve. State the optimisations used, where and in what order they were applied, and the set of variables live (under both forms of liveness) on entry to the optimised version of `p`. [6 marks]

14 Topical Issues

A company making social networking systems is seeking to create a location-aware application to alert users when their friends are less than 50m away. The application should work in three public environments: parks, city streets and indoor spaces such as shopping centres. It will run on smartphones, which may be assumed to have radios for 2G/3G cellular networks, WiFi, Bluetooth, and GPS. These radios can be assumed to be permanently enabled.

(*a*) Characterise the three public environments in terms of the opportunities and challenges they present for location tracking. [6 marks]

(*b*) The company will use fingerprinting with 2G/3G cellular signals only in the three public environments. They will use the RSSI value reported for each cellular tower, and the nearest neighbour technique to match vectors of RSSI values.

(*i*) Discuss how appropriate the fingerprinting technique is in this context. [4 marks]

(*ii*) To compare RSSI vectors they intend to use Manhattan distance rather than Euclidean distance, i.e. for two vectors $\mathbf{x}$ and $\mathbf{y}$,

$$D_{Manhattan} = \sum_i |x_i - y_i|,$$

where i spans every component of the vectors. By comparing the vector (2,2,2) with (8,8,8), (12,9,1), and (16,0,0), using both Euclidean and Manhattan distance, discuss the advantages and disadvantages of this approach. [7 marks]

(*iii*) The database of fingerprints will be large and continually updated. By analogy with A-GPS, suggest how to distribute the map to smartphones efficiently. [3 marks]

END OF PAPER