COMPUTER SCIENCE TRIPOS Part II

Tuesday 5 June 2012 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 Advanced Graphics

(*a*) *NURBS* and *subdivision* are alternative methods for representing surfaces in three dimensions.

(*i*) Compare and contrast *NURBS* and *subdivision* for representing surfaces in three dimensions. [5 marks]

(*ii*) Suggest why the animation industry favours subdivision and the CAD industry favours NURBS. [2 marks]

(*b*) Chaikin's corner-cutting curve subdivision method is based on the quadratic uniform B-spline.

(*i*) Describe Chaikin's method. [3 marks]

(*ii*) Describe the quadratic uniform B-spline for representing curves. [4 marks]

(*iii*) Describe how Chaikin's method for curves is extended and generalised to produce a subdivision method for surfaces that is able to cope with polygons with any number of sides and to cope with vertices with any number of incident edges. [6 marks]

2 Artificial Intelligence II

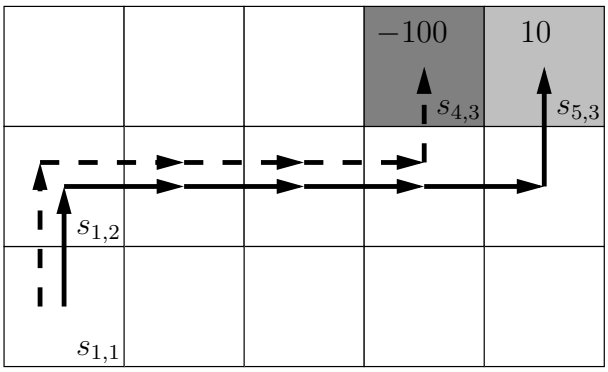
A *reinforcement learning problem* has states $\{s_1, \ldots, s_n\}$, actions $\{a_1, \ldots, a_m\}$, reward function $R(s, a)$ and next state function $S(s, a)$.

(a) Give a general definition of a *policy* for such a problem. [1 mark]

(b) Give a general definition of the *discounted cumulative reward* and the corresponding *optimal policy* for such a problem. [5 marks]

(c) Give an expression for the optimal policy in terms of R, S and the discounted cumulative reward, and show how this can be modified to produce the *Q-learning algorithm*. [7 marks]

In a simple reinforcement learning problem, states are positions on a grid and actions are `up` and `right`. The only way an agent can receive a reward is by moving into one of two special positions, one of which has a reward of 10 and the other of -100.



Here, states are labelled by their grid coordinates. A possible sequence of actions (sequence 1) is shown by solid arrows, ending with a reward of 10 being received, and another (sequence 2) by dashed arrows ending with a reward of -100.

(d) Assume that all Q values are initialised at 0.

(i) Explain how the Q values are altered if sequence 1 is used *twice* in succession by the Q-learning algorithm. [4 marks]

(ii) Explain what further changes occur to the Q values if sequence 2 is then used *once* by the Q-learning algorithm. [3 marks]

3 Bioinformatics

(*a*) Considerable recent Bioinformatics research has focused on *phylogenetics*.

(*i*) What is the motivation for this work? [1 mark]

(*ii*) Describe with the aid of examples *two* different techniques for phylogeny. In each case discuss the issues of complexity and performance. [4 marks each]

(*b*) Considerable recent Bioinformatics research has focused on *structure prediction from sequence data*.

(*i*) Describe how you would build a hidden Markov model (HMM) to identify membrane segments in aminoacid sequences. [6 marks]

(*ii*) How you would assess the sensitivity and specificity performance of your HMM? [5 marks]

4 Business Studies

(*a*) Distinguish between a profit and loss statement and cash flow statement. [5 marks]

(*b*) A company is proposing to build a low cost single board computer that will sell for £30 direct from its web page. The bill of materials (BoM) costs for components are £15/unit, while manufacturing and other costs are estimated as 33% of sale price. Components can only be bought in lots of 10,000 at a time and must be paid for 1 month before first use. The company estimates sales for the first six months as a ramp for 0 in the first month increasing by 1000/month to 5000 units in month 6.

(*i*) Draw up a cash flow estimate for the first six months of operation. Ignore VAT, bank and other charges. [5 marks]

(*ii*) How much working capital will be required? [5 marks]

(*iii*) In month 5 the company is offered the opportunity to sell 50,000 units but at a price of £25/unit. Should the company take this opportunity? Justify your answer. [5 marks]

5 Comparative Architectures

(*a*) How can specialising a processor design for a specific application domain help to reduce its power consumption? [7 marks]

(*b*) For many applications chip-multiprocessors consume less power when compared to a superscalar uniprocessor design of equal performance. Why is this the case? [7 marks]

(*c*) Why does a vector processor offer a particularly energy efficient solution to execute some types of program? [6 marks]

6 Denotational Semantics

(*a*) If D and D' are domains, explain what is the *function domain* $D \to D'$; give its partial order and least element, and explain how least upper bounds of chains are calculated in it. [4 marks]

(*b*) An element d of a domain D is said to be *isolated* if for all countable chains $x_0 \sqsubseteq x_1 \sqsubseteq x_2 \sqsubseteq \ldots$ in D with $d \sqsubseteq \bigsqcup_{n \geq 0} x_n$, there exists $i \geq 0$ with $d \sqsubseteq x_i$. We write $K(D)$ for the subset of isolated elements.

Given domains D and D' and elements $d \in D$ and $d' \in D'$, let $[d, d'] : D \to D'$ be the function mapping each $x \in D$ to d' if $d \sqsubseteq x$ and to $\bot$ otherwise.

(*i*) Prove that $[d, d']$ is monotone. [2 marks]

(*ii*) Prove that if $f : D \to D'$ is monotone, then $[d, d'] \sqsubseteq f$ if and only if $d' \sqsubseteq f(d)$. [2 marks]

(*iii*) Prove that if $d \in K(D)$, then $[d, d']$ is an element of the function domain $D \to D'$. [3 marks]

(*iv*) Prove that if both $d \in K(D)$ and $d' \in K(D')$, then $[d, d']$ is an isolated element of the function domain $D \to D'$. [3 marks]

(*v*) Now suppose that every element of D is the least upper bound of some countable chain of isolated elements and the same is true for D'. Show that each element f of the function domain $D \to D'$ is the least upper bound of the subset $F \stackrel{\text{def}}{=} \{[d, d'] \mid d \in K(D) \;\&\; d' \in K(D') \;\&\; d' \sqsubseteq f(d)\}$. [6 marks]

7 Hoare Logic

In this question we consider a semantics of FOR-commands in which

$\texttt{FOR}\ V\texttt{:=}E_1\ \texttt{UNTIL}\ E_2\ \texttt{DO}\ C$

is defined to be equivalent to

$V\texttt{:=}E_1\texttt{;}\ \texttt{WHILE}\ V \leq E_2\ \texttt{DO}\ \texttt{(}C\texttt{;}\ V\texttt{:=}V+1\texttt{)}$

(*a*) How does this semantics of FOR-commands differ from the one given in the lectures? [4 marks]

(*b*) The following FOR-rule is similar to one proposed by John Wickerson:

$$\frac{\vdash P \Rightarrow R[E_1/V] \quad \vdash R \wedge V{>}E_2 \Rightarrow Q \quad \vdash \{R \wedge V \leq E_2\}\ C\ \{R[V{+}1/V]\}}{\vdash \{P\}\ \texttt{FOR}\ V\texttt{:=}E_1\ \texttt{UNTIL}\ E_2\ \texttt{DO}\ C\ \{Q\}}$$

Assuming the semantics of FOR-commands given above, derive this Wickerson-style FOR-rule from the standard axioms and rules of Hoare logic. [10 marks]

(*c*) Is the FOR-axiom:

$\vdash \{P \wedge E_2{<}E_1\}\ \texttt{FOR}\ V\texttt{:=}E_1\ \texttt{UNTIL}\ E_2\ \texttt{DO}\ C\ \{P\}$

sound with the semantics given above? Justify your answer either with a proof of this axiom, or with a counterexample. [6 marks]

8 Information Theory and Coding

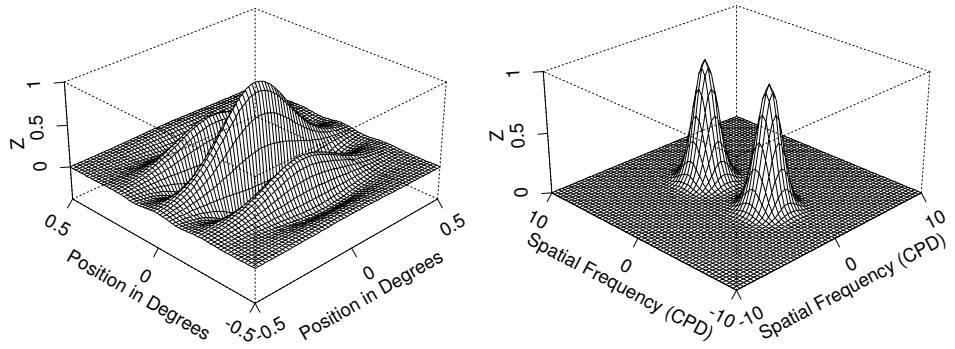
(a) Consider two independent integer-valued random variables, X and Y. Variable X takes on only the values of the eight integers $\{1, 2, \ldots, 8\}$ and does so with uniform probability. Variable Y may take the value of *any* positive integer k, with probabilities $P\{Y = k\} = 2^{-k},\ k = 1, 2, 3, \ldots$.

(i) Which random variable has greater uncertainty? Calculate both entropies $H(X)$ and $H(Y)$. [3 marks]

(ii) What is the joint entropy $H(X, Y)$ of these random variables, and what is their mutual information $I(X; Y)$? [2 marks]

(b) What is the maximum possible entropy H of an alphabet consisting of N different letters? In such a maximum entropy alphabet, what is the probability of its most likely letter? What is the probability of its least likely letter? Why are fixed length codes inefficient for alphabets whose letters are not equiprobable? Discuss this in relation to Morse Code. [5 marks]

(c) Explain why the real-part of a 2D Gabor wavelet has a 2D Fourier transform with two peaks, not just one, as shown in the right panel of the figure below. [4 marks]



(d) Show that the set of all Gabor wavelets is closed under convolution, *i.e.* that the convolution of any two Gabor wavelets is just another Gabor wavelet. [*Hint:* This property relates to the fact that these wavelets are also closed under multiplication, and that they are also self-Fourier. You may address this question for just 1D wavelets if you wish.] [3 marks]

(e) Show that the family of sinc functions used in the Nyquist Sampling Theorem,

$$\mathrm{sinc}(x) = \frac{\sin(\lambda x)}{\lambda x}$$

is closed under convolution. Show further that when two different sinc functions are convolved together, the result is simply whichever one of them had the lower frequency, *i.e.* the smaller λ. [3 marks]

9 Natural Language Processing

(*a*) Define the terms *morpheme, affix, circumfix* and *derivational morphology* and give examples. [4 marks]

(*b*) What is a finite state transducer (FST), and what is it used for in computational linguistics? How does it differ from a finite state automaton? [3 marks]

(*c*) Draw an FST which recognises the affix *-ly* associated with regular adverbs in English. Demonstrate that your FST correctly handles cases such as *bright* $\rightarrow$ *brightly, simple* $\rightarrow$ *simply, silly* $\rightarrow$ *sillily* and *terrific* $\rightarrow$ *terrifically.* [9 marks]

(*d*) In terms of linguistic phenomena, which aspects of the morphology of adverbs does your FST fail to handle correctly? Could these problems be fixed with a more involved FST, or are they general problems with FST? [4 marks]

10 Optimising Compilers

(*a*) Annotate the uses of a given variable with '*D*' (definition), '*R*' (reference) and '*U*' (undefinition—e.g. entering or leaving scope). Explain why three pairs of such uses may be regarded as *anomalous* data flow, giving brief example programs illustrating how each of these anomalies can represent a programmer error. Your answer should make use of the concept of path in a dataflow graph. [3 marks]

(*b*) Give a single program containing all three of the above anomalies and justify it as *not* containing any programmer error. [3 marks]

(*c*) By analogy with "on all paths" and "on some path" in dataflow analysis refine your definition of dataflow anomaly in part (*a*) into that of 'must-anomaly' and 'may-anomaly' indicating which, if either, of these corresponds to the previous definition. How many must-anomalies does your program in part (*b*) contain? [3 marks]

(*d*) Give *two* sets of dataflow equations, in the style used to define "live variables" or "available expressions", which each calculate a set of variables for every program point. Explain how these can respectively be used to issue compiler warning messages of the form "*variable* '`x`' *may be read before being set*" and "*variable* '`x`' *is definitely read before being set*". For both forms, state the initialisation of these sets of variables to be used when solving the dataflow equations. [7 marks]

(*e*) Now suppose the language supports indirect assignment to address-taken variables as in C. Explain briefly the refinements necessary to the above analyses. [2 marks]

(*f*) The wording ('is definitely read') of the second message in part (*d*) may be criticised when considering the program

```
{ int x,y=0; while (...) y=f(y); print x; }
```

Comment briefly on this claim. [2 marks]

11 Principles of Communications

(*a*) Feedback-based congestion control is employed by the Internet's transport protocol, TCP (the Transmission Control Protocol), using a window-based approach. An alternative is to adjust the sender's *rate* explicitly.

(*i*) What feedback signals are used to adjust the window-based scheme? [4 marks]

(*ii*) What additionally has to be measured for the rate-based approach? [2 marks]

(*iii*) In feedback-based congestion control schemes, the signal is sometimes referred to as a *price*. Explain, perhaps with reference to a diagram, why this is so. [3 marks]

(*b*) Round Robin Schedulers provide what is known as *Fair Queueing*.

(*i*) What is a general formulation of the max-min fairness property? [4 marks]

(*ii*) Where do the overheads arise in implementing Fair Queueing in an IP router? [4 marks]

(*iii*) In what circumstances might a FIFO (first-in first-out) queue approximate a fair queue? [3 marks]

12 Security II

The RSA cryptosystem can be tuned to make the workload asymmetric: with $d = 3$, encryption (cubing modulo n) becomes very cheap and almost all the computational expense shifts to decryption (extracting cubic roots modulo n).

The following public-key protocol uses the above property to allow two principals A and B to establish a common secret key N_b (invented by B) without incurring a high computational load, thanks to the help of a server S who computes all the cubic roots in the protocol. Attackers are assumed to be able to overhear, but not alter, the messages between A, B and S.

$$\begin{array}{ll} A \to S: & B, N_a^3 \bmod n. \\ S \to B: & A. \\ B \to S: & A, N_b^3 \bmod n. \\ S \to A: & B, N_a \oplus N_b. \end{array}$$

(a) What is the purpose of N_a? [3 marks]

(b) Describe in detail a protocol attack that will allow two colluding attackers C and D to recover N_b. Assume that S is stateless. [7 marks]

(c) Stop the attack you described in (b) by making S stateful. [3 marks]

(d) Describe in detail a more sophisticated protocol attack whereby the colluding attackers will recover N_b even if S adopts the precaution you described in (c). [4 marks]

(e) Fix the protocol to defeat the attack you described in (d). [3 marks]

13 Temporal Logic and Model Checking

In the following program, called `INC`, `(*)` is a Boolean expression that evaluates non-deterministically to either true or false each time it is evaluated (different evaluations may yield different results).

```
N := 1;
WHILE (*) DO N := N+1;
N := 0;
```

(*a*) Devise and carefully describe a state space and transition relation to model this program. [8 marks]

(*b*) Devise temporal logic formulae that express properties (*i*), (*ii*) and (*iii*) below. In each case state which temporal logic you are using and explain whether the property is true in your model.

(*i*) Every execution of `INC` terminates. [4 marks]

(*ii*) Some execution of `INC` terminates. [4 marks]

(*iii*) If `INC` terminates, then `N` $= 0$ holds in the terminating state. [4 marks]

14 Topical Issues

(*a*) Describe the operating principles of RFID systems based on capacitive, inductive and backscatter coupling. Give typical operating ranges and at least one example application for each. [9 marks]

(*b*) Electronic passports contain RFID chips to permit electronic transfer of biometric data to a reader. The systems in use today use remote-coupling with an operating frequency of 13.56 MHz. The full communication protocol varies, but every transmission from the passport is prefaced with a numeric identifier (UID).

(*i*) The first implementations used a static UID unique to each passport, whilst later implementations generated a new pseudo-random UID for each round of communications. Outline the risks and practicalities of each approach. [3 marks]

(*ii*) Biometric data are usually encrypted when sent between reader and tag using a protocol known as Basic Access Control (BAC). The shared session key is generated solely from the owner's passport number (9 digits), passport expiry date and date of birth. These data must be read *optically* by swiping the passport through a desktop device before proceeding. Comment on the security of this system and the choice of 13.56 MHz RFID in such a context. [4 marks]

(*iii*) Most authorities now line the passport sleeve with metal foil. Explain how this increases security and discuss the extent to which it does so. [4 marks]

END OF PAPER