COMPUTER SCIENCE TRIPOS Part II

Tuesday 1 June 2010 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 Additional Topics

The IT industry is presently interested in minimising the power consumption associated with technology.

(*a*) Explain what is meant by the term *energy-proportional computing* in relation to server farms and show how it might be achieved. [7 marks]

(*b*) Minimising the power consumption of desktop and mobile devices has traditionally been done by adapting system components such as the CPU to support multiple power states. An alternative is to incorporate two or more chips designed to give the same result but with different power/performance characteristics and to switch between them. For example, a system may switch between a high power and a low power CPU. Compare and contrast these two approaches. [6 marks]

(*c*) Engineering for energy efficiency is a common task for mobile computing devices such as laptops. By considering the typical work load profiles of laptops and servers, and the environments in which they operate, discuss whether the same solutions developed for laptops are likely to be applicable to servers. [7 marks]

2 Advanced Graphics

(a) For the knot vector $[0, 0, 0, 1, 1, 1]$:

(i) Derive all of the B-spline basis functions for the quadratic case ($k = 3$). [5 marks]

(ii) Place three control points, $P_1, \ldots, P_3$ in an equilateral triangle. Draw the B-spline curve, $P(t)$, defined by the control points and the basis functions derived from the knot vector. [2 marks]

(iii) Identify and label appropriately the locations on the curve in part (a)(ii) at which the parameter, t, is an integer. [1 mark]

(iv) Identify and express as a function of the control point locations the points at which the curve in part (a)(ii) touches the triangle defined by the control points, $P_1, \ldots, P_3$. [1 mark]

(b) For the uniform knot vector $[0, 1, 2, 3, 4, 5, 6]$:

(i) Sketch all of the quadratic ($k = 3$) B-spline basis functions. [2 marks]

(ii) Place four control points, $Q_1, \ldots, Q_4$ in a square. Draw the B-spline curve, $Q(t)$, defined by the control points and the basis functions derived from the knot vector. [2 marks]

(iii) Identify and label appropriately the locations on the curve in part (b)(ii) at which the parameter, t, is an integer. [1 mark]

(iv) Identify and express as a function of the control point locations the points at which the curve in part (b)(ii) touches the square defined by the control points, $Q_1, \ldots, Q_4$. [1 mark]

(v) Give the continuity of the curve at $t = 3$. [1 mark]

(c) Describe the corner-cutting curve subdivision algorithm (Chaikin algorithm) that, in the limit, produces the same curve as the uniform quadratic B-spline. [4 marks]

3 Advanced Systems Topics

(*a*) (*i*) Explain the collision detection mechanism applied in standard wired medium access control associated with CSMA and indicate why this might be unsuitable for wireless networks. [2 marks]

(*ii*) Describe how CSMA/CA (Collision Avoidance) works and explain its limitations. [2 marks]

(*iii*) Illustrate how MACAW works and indicate its limitations on the exposed terminal problem through an example. [3 marks]

(*iv*) Would the deployment of MACAW be needed in an installation of a MAC protocol for a single base station that communicates with some mobile terminals? Why? [3 marks]

(*b*) You are called upon, as expert, to design and deploy a wireless sensor network to monitor environmental factors in a 500 m^2 area within a forest.

(*i*) Describe the network architecture you would deploy to collect sensed data in the forest. [2 marks]

(*ii*) Briefly illustrate **two** different MAC layer protocols for sensor networks and then indicate which one you would employ for your network and why. [4 marks]

(*iii*) Explain the salient characteristics of Directed Diffusion (DD). Describe the process through which DD is able to reconfigure when sensor nodes fail in the network. [4 marks]

4 Artificial Intelligence II

Professor Elbow-Patch is not the man he used to be, and in particular has a tendency to fall over for no apparent reason. This problem is made worse if he has drunk port with his dinner. He almost always drinks port on a Sunday, and if he drinks on any given day he is unlikely—for the sake of his long-suffering liver—to drink port on the following day. However, if he does not drink on a given day then he is very likely to succumb to temptation on the following day.

The probability that he falls over after drinking is $\Pr(\text{fall}|\text{drank}) = 0.7$. The probability that he falls over when he has not drunk is $\Pr(\text{fall}|\neg\text{drank}) = 0.1$. He drinks on a Sunday with probability 0.9. If he has not drunk on a given day then the probability that he drinks the following day is $\Pr(\text{drink today}|\neg\text{drank yesterday}) = 0.8$. If he has drunk on a given day then the probability that he drinks the following day is $\Pr(\text{drink today}|\text{drank yesterday}) = 0.1$.

(*a*) Explain how this problem can be represented as a *hidden Markov model*. What assumptions are required? [4 marks]

(*b*) Denoting observations at time i by E_i and states at time i by S_i give a derivation of the *filtering algorithm* for computing $\Pr(S_t|E_1, \ldots, E_t)$. [8 marks]

(*c*) You observe the Professor on Sunday, Monday and Tuesday and notice that he does not fall over at all. Use the filtering algorithm to compute the probability that he drank port on Tuesday. [8 marks]

5 Bioinformatics

(*a*) Discuss the space–time complexity of dynamic programming algorithms in sequence alignment. [7 marks]

(*b*) Discuss with one example how to score a multiple sequence alignment. [5 marks]

(*c*) Describe the use of principal component analysis in microarray data analysis. [8 marks]

6 Business Studies

A certain company decides to make a new educational computer. They agree terms with an OEM to supply units at a cost of £10/unit delivered, payable by LoC on shipment. They predict sales per quarter in units as follows:

Quarter	Q1	Q2	Q3	Q4
Sales (units)	200	2000	4000	8000

The units are sold to a distributor as they are delivered at a price of £30 wholesale, for a target price of £49.99 retail.

The company employs five staff at a monthly salary of £2000 each, and has overheads equivalent to salary. In addition there is a marketing spend of £10,000 per quarter.

(*a*) Draw up a quarterly profit and loss statement. [5 marks]

(*b*) Draw up a quarterly cash flow projection for this project, stating assumptions. How much working capital is required? [5 marks]

(*c*) Draw up a balance sheet at the end of Q4. [5 marks]

(*d*) Comment on how the working capital might be funded. [5 marks]

7 Comparative Architectures

(*a*) Why is a shared second-level (L2) cache typically divided into multiple banks (banked) in a chip multiprocessor? [3 marks]

(*b*) In what situation might a shared second-level cache offer a performance advantage over a memory hierarchy for a chip multiprocessor with private L2 caches? [4 marks]

(*c*) A cache controller in a chip multiprocessor snoops the bus and observes a transaction that refers to a block that its cache contains. The block is held in State M (Modified). The bus transaction has been generated by a processor wishing to read the block. Assuming a MSI (write-back invalidate) cache coherence protocol, what actions will be taken by the cache controller? [6 marks]

(*d*) How does adopting an inclusion policy simplify the implementation of a cache coherence mechanism in a chip multiprocessor with private L1 and L2 caches? [4 marks]

(*e*) How might multiple buses be exploited to enable a greater number of processors to be supported by a snoopy cache coherence protocol? [3 marks]

8 Denotational Semantics

(*a*) For a domain D, recall that by Tarski's Fixed-Point Theorem every continuous function $f \in (D \to D)$ has a least pre-fixed point $\mathit{fix}(f) \in D$.

Prove that the function $\mathit{fix} : (D \to D) \to D$ is continuous. [10 marks]

(*b*) For a partially ordered set $(P, \sqsubseteq)$, let $\big(\mathrm{Ch}(P), \sqsubseteq_{\mathrm{ptw}}\big)$ be the partially ordered set of chains in P ordered pointwise. That is,

$$\mathrm{Ch}(P) \stackrel{\mathrm{def}}{=} \left\{ x = \{x_n\}_{n \in \mathbb{N}} \;\middle|\; \text{for all } i \leq j \text{ in } \mathbb{N},\ x_i \sqsubseteq x_j \text{ in } P \right\}$$

and

$$x \sqsubseteq_{\mathrm{ptw}} x' \iff_{\mathrm{def}} x_n \sqsubseteq x'_n \text{ for all } n \in \mathbb{N}$$

Show that if P is a domain then so is $\mathrm{Ch}(P)$. [10 marks]

9 Digital Communication II

(*a*) Packet size has a marked influence on network performance.

(*i*) What is the motivation for the minimum packet size on Ethernet? [3 marks]

(*ii*) What is the rationale for the maximum packet size on Ethernet? [2 marks]

(*b*) The max–min fair share criterion for allocation m_n of resources to a set of N flows, with respective demands x_n, for a resource of capacity C, can be computed using the following equations:

$$m_n = min(x_n, M_n), for\ 1 <= n <= N$$

and

$$M_n = \frac{C - \sum_{i=1}^{n-1} m_i}{N - n + 1}$$

Explain, perhaps with the use of an example, how this criterion operates to mitigate between over- and under-demands. [5 marks]

(*c*) Packet Schedulers in switches and routers provide isolation, and fairness, but are subject to implementation efficiency considerations which lead to inaccuracy in how fair they are. Discuss the impact of packet sizes on the accuracy of resource allocation compared with an idealised scheduler, making reference to a typical practical example such as Weighted Round Robin, or Deficit Round Robin. [10 marks]

10 Human–Computer Interaction

You have been asked to design a programming environment for use by teachers and school children in rural India, that will run on a mobile phone.

(*a*) Using the Cognitive Dimensions of Notations, describe **four** different usability issues that are likely to arise specifically when a user is creating and modifying programs on a mobile phone. [2 marks each]

(*b*) Describe techniques that you would use to research the requirements of the teachers and children. [4 marks]

(*c*) What factors are likely to be important in assessing whether the environment will be of value to *school children*? How would you measure and evaluate those factors? [4 marks]

(*d*) What factors are likely to be important in assessing whether the environment will be of value to *teachers*? How would you measure and evaluate those factors? [4 marks]

11 Information Theory and Coding

(a) Let X and Y be two discrete random variables whose respective sets of possible outcomes $\{x\}$ and $\{y\}$ are described by probability distributions $p(x)$ and $p(y)$, and by a joint probability distribution $p(x, y)$.

(i) Give an expression for the *mutual information* $I(X;Y)$ between X and Y, using only the probability distributions $p(x)$, $p(y)$, and $p(x, y)$. [2 marks]

(ii) In case X and Y are independent random variables, what becomes of their mutual information, and why? [1 mark]

(iii) Let the marginal entropy of random variable X be $H(X)$, and suppose that the two random variables X and Y are perfectly correlated with each other. In that case, prove that $I(X;Y) = H(X)$. [2 marks]

(iv) What is $I(X;X)$, the mutual information of a random variable with itself, in terms of $H(X)$? [1 mark]

(b) Prove that the information measure is additive: that the information gained from observing the combination of N independent events, whose probabilities are p_i for $i = 1....N$, is the *sum* of the information gained from observing each one of these events separately and in any order. [3 marks]

(c) An invertible transform generates projection coefficients by integrating the product of a signal onto each of a family of functions. In a reverse process, expansion coefficients can be used on those same functions to reproduce the signal. If the functions in question happen to form an orthonormal set, what is the consequence for the projection coefficients and the expansion coefficients? What penalty must be paid in the absence of orthogonality? Name one such penalised transform. [3 marks]

(d) In the Information Diagram (a plane whose axes are time and frequency), why does the Gabor–Heisenberg–Weyl *Uncertainty Principle* imply that information is *quantised* even in continuous signals – *i.e.,* that a signal's information content consists of a countable, limited number of independent quanta? [3 marks]

(e) Define the Kolmogorov algorithmic complexity K of a string of data. For a string of length N bits, how large might its Minimal Description Length be, and why? Comment on how, or whether, you can know that you have truly determined the Minimal Description Length for a set of data. Give a reasonable estimate of the Kolmogorov complexity K of a fractal, and explain why it is reasonable. [5 marks]

12 Optimising Compilers

A new programming language follows an "evaluate only on need" strategy. The consequence is, for example, that in a program fragment

```
int x = SomeExpression;
if (SomeBoolean) print x;
if (x == 0) x = SomeOtherExpression;
```

the given (potentially complicated) expression is not evaluated on the line that declares `x`, but only gets evaluated if and when `x` is used, as in the print statement on line 2. On line 3 the value of `x` is certainly needed, so if the expression had not been evaluated earlier it must be there.

Optimising compilation normally improves performance of compiled code based on transformations that are "safe" in some sense. Propose forms of analysis and hence optimisation relevant in this case in the circumstances:

(*a*) It is safe but undesirable to evaluate an expression even if its value will not subsequently be used, but it must not be evaluated a second time. [8 marks]

(*b*) It is safe to evaluate an expression repeatedly if is evaluated at all, but if the program would not use the value it must not be computed at all. [7 marks]

(*c*) Extra executable logic has to be put in the program to ensure that each expression is evaluated exactly once if its value is needed, but not at all otherwise. It is desirable to minimise the amount of this extra logic, preserving semantics exactly. [5 marks]

You need only consider the case of optimisation of a single procedure at a time.

13 Security

Woo & Lam proposed a protocol that would enable a client A to log on to a server B using an authentication service S.

$$\begin{array}{l} \mathrm{A} \rightarrow \mathrm{B}: \ A \\ \mathrm{B} \rightarrow \mathrm{A}: \ N_B \\ \mathrm{A} \rightarrow \mathrm{B}: \ \{N_B\}_{KAS} \\ \mathrm{B} \rightarrow \mathrm{S}: \ \{A, \{N_B\}_{KAS}\}_{KBS} \\ \mathrm{S} \rightarrow \mathrm{B}: \ \{N_B\}_{KBS} \end{array}$$

(*a*) Explain the protocol notation. [4 marks]

(*b*) Explain why the protocol is insecure. [12 marks]

(*c*) How should it be fixed? [4 marks]

14 Specification and Verification I

(*a*) Explain carefully the meaning of a Hoare specification $\{P\}C\{Q\}$ and describe the nature of P, C and Q. [4 marks]

(*b*) How can terms in higher-order logic be used to represent P, C and Q? Give the types of the terms. [4 marks]

(*c*) Describe the meaning of the *weakest liberal precondition* (wlp) and the *strongest postcondition* (sp). [4 marks]

(*d*) Define wlp and sp in higher-order logic. What are the types of wlp and sp? [4 marks]

(*e*) Explain the relationships between Hoare specifications, weakest liberal preconditions and strongest postconditions. What is the importance of this for automating program verification? [4 marks]

15 Specification and Verification II

(*a*) Describe the kinds of property that can be expressed in LTL, CTL, CTL*, ITL and PSL. [6 marks]

(*b*) What kind of properties are expressed by SEREs in PSL? Give an example of a SERE and explain its meaning. [4 marks]

(*c*) It is sometimes said that engineers prefer LTL to CTL. Why might this be the case? [4 marks]

(*d*) Are there properties that can be expressed in CTL that cannot be expressed in LTL? Justify your answer. [2 marks]

(*e*) Are there properties that can be expressed in CTL that cannot be expressed in PSL? Justify your answer. [2 marks]

(*f*) Are there properties that can be expressed in PSL that cannot be expressed in LTL? Justify your answer. [2 marks]

END OF PAPER