COMPUTER SCIENCE TRIPOS Part II

Thursday 7 June 2007 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

STATIONERY REQUIREMENTS
Script paper
Blue cover sheets
Tags

SPECIAL REQUIREMENTS
None

1 Advanced Systems Topics

(*a*) Microkernel operating systems aim to address perceived modularity and reliability issues in traditional "monolithic" operating systems.

(*i*) Describe the typical architecture of a microkernel operating system such as Mach 3.0, including its use of address spaces, processor privilege levels and communication channels compared with a traditional operating system structure. [6 marks]

(*ii*) How does this structure affect the system performance? [4 marks]

(*b*) Virtual machine monitors existed before microkernels and have experienced a recent resurgence of interest; using examples, describe some of the motivations for a virtual machine monitor approach. [4 marks]

(*c*) Compare and contrast the means used to support pre-existing applications in the microkernel and virtual machine worlds. [6 marks]

2 Natural Language Processing

(*a*) There are several different types of ambiguity in natural language. For *each* of the following cases, briefly describe how ambiguity may arise, illustrating your answer with examples:

(*i*) morphological ambiguity;

(*ii*) lexical ambiguity;

(*iii*) syntactic ambiguity;

(*iv*) ambiguity in rhetorical relations (discourse relations).

[12 marks total]

(*b*) Explain what is meant by "packing" in parsing and discuss its relevance to the treatment of ambiguity. [8 marks]

3 VLSI Design

(*a*) What is meant by *dual-rail* logic? When is it used? [4 marks]

(*b*) Describe the purpose and operation of a Muller C element. [3 marks]

(*c*) Sketch a transistor-level circuit for a 3-input C element in CMOS, including an inverted output. [3 marks]

(*d*) Consider the design of a one-bit full-adder in asynchronous dual-rail logic. There will be three input pairs: (P_0, P_1), (Q_0, Q_1) and (R_0, R_1), and two output pairs: (S_0, S_1) for the high-order bit of the sum and (T_0, T_1) for the low-order bit. All inputs should become valid before any outputs become valid, and then all the outputs should remain valid until all the inputs are clear again. Sketch a circuit diagram for the adder using inverters, NAND gates, NOR gates and C elements. [10 marks]

4 Digital Communication II

(*a*) Many distributed algorithms for sharing resources on the time scale of a round-trip time exist.

(*i*) Describe the p-persistent CSMA/CD algorithm used on Ethernets to manage capacity and contention for a shared resource. [5 marks]

(*ii*) Describe a difference from (*i*) arising from the hidden- and exposed-terminal problems in shared-media wireless media networks, and how this leads to CSMA/CA. [5 marks]

(*b*) Discuss how the use of TCP's congestion avoidance and control algorithm can be seen as similar to the contention schemes used in Ethernets (wired or wireless). [5 marks]

(*c*) What is the key parameter distinguishing between CSMA/CD and TCP, and why does this make a difference in the different schemes for avoiding congestion and contention on shared media? [5 marks]

5 Distributed Systems

(*a*) When distributed systems are designed and engineered, certain fundamental properties have to be taken into account, including:

1. concurrent execution of components
2. independent failure modes
3. communication delay
4. no global time

Give *three* examples of the implications of these properties (separately or in combination) on the engineering of large-scale, widely distributed systems.

[9 marks]

(*b*) (*i*) Define role-based access control (RBAC).

(*ii*) Outline how RBAC could be used for a national healthcare system comprising many administration domains such as primary care practices, hospitals, specialist clinics, etc. Principals may, from time to time, work in domains other than their home domain, and must be authorised to do so.

(*iii*) A national Electronic Health Record (EHR) service must be accessible from all domains. It is required by law that access control policy should be able to capture exclusion of principals and relationships between them. How could this requirement be met in an RBAC design?

[11 marks]

6 Computer Vision

(*a*) Briefly define each of the following concepts as it relates to vision:

(*i*) active contour; [2 marks]

(*ii*) Hadamard's criteria for well-posed problems; [2 marks]

(*iii*) functional streaming; [2 marks]

(*iv*) reflectance map; [2 marks]

(*v*) Bayesian prior. [2 marks]

(*b*) What is accomplished by the lateral signal flows within both plexiform layers of the mammalian retina, in terms of spatial and temporal image processing and coding? [3 marks]

(*c*) Give finite difference operators that could be applied to 1-dimensional discrete data (such as a row of pixels) in order to approximate the 1st and 2nd derivatives, $\frac{d}{dx}$ and $\frac{d^2}{dx^2}$. How would your finite difference operators actually be applied to the row of pixels? What is the benefit of using a 2nd finite difference (or derivative) instead of a 1st finite difference (or derivative) for edge detection? [3 marks]

(*d*) Explain the formal mathematical similarity between the "eigenface" representation for face recognition, and an ordinary Fourier transform, in the following respects:

(*i*) Why are they both called linear transforms, and what is the "inner product" operation in each case? [1 mark]

(*ii*) What is a projection coefficient and an expansion coefficient in each case? [1 mark]

(*iii*) What is the orthogonal basis in each case, and what is meant by orthogonality? [1 mark]

(*iv*) Finally, contrast the two in terms of the use of a data-dependent or a data-independent (universal) expansion basis. [1 mark]

7 Advanced Graphics

(*a*) Outline a method that could be used for modelling water and other fluids. [6 marks]

(*b*) (*i*) Show how to find the first intersection between a ray and a finite-length, open-ended cone, of unit slope, centred at the origin, aligned along the y-axis, for which both ends of the finite-length are on the positive y-axis (i.e. $0 < y_{min} < y_{max}$). [6 marks]

(*ii*) Extend this to cope with a closed cone (i.e. the same cone, but with end caps). Take care to consider any special cases. [4 marks]

(*iii*) Extend this further to give the normal vector at the intersection point. Take care to consider all cases. [4 marks]

8 Specification and Verification II

(*a*) Given a predicate $\mathcal{B}$ specifying a set of initial states and a transition relation $\mathcal{R}$, define the set of reachable states. [4 marks]

(*b*) Explain the difference between *Linear Temporal Logic* and *Interval Temporal Logic*. [4 marks]

(*c*) (*i*) What is *model checking*? [2 marks]

(*ii*) Explain the key idea of symbolic model checking. [2 marks]

(*d*) (*i*) What is a SERE in PSL? [2 marks]

(*ii*) Give an example of a PSL formula involving a SERE and explain its meaning. [2 marks]

(*e*) (*i*) What is the difference between *dynamic verification* and *static verification*? [2 marks]

(*ii*) What kind of tools are used with each sort of verification? [2 marks]

9 Artificial Intelligence II

An agent exists within an environment in which it can perform actions to move between states. On executing any action it moves to a new state and receives a reward. The agent aims to explore its environment in such a way as to learn which action to perform in any given state so as in some sense to maximise the accumulated reward it receives over time.

(*a*) Give a detailed definition of a *deterministic Markov decision process* within the stated framework. [4 marks]

(*b*) Give a general definition of a *policy*, of the *discounted cumulative reward*, and of the *optimum policy* within this framework. [4 marks]

(*c*) Give a detailed derivation of the *Q-learning* algorithm for learning the optimum policy. [8 marks]

(*d*) Explain why it is necessary to trade-off *exploration* against *exploitation* when applying Q-learning, and explain one way in which this can be achieved in practice. [4 marks]

10 Digital Signal Processing

(*a*) Write an efficient microcontroller program (pseudocode) that outputs a continuous sine wave of frequency $f = 440$ Hz with values y_n in the range -1 to 1 at a sampling frequency $f_\mathrm{s} = 32$ kHz. The programming language you have available lacks complex-number arithmetic, the runtime environment offers only basic floating-point arithmetic (i.e., no trigonometric functions), addition is much faster than multiplication, and there is insufficient memory to store a precomputed waveform. [10 marks]

(*b*) The discrete sequence $y_n = \cos(2\pi n f_1/f_\mathrm{s}) + A \cdot \cos(2\pi n f_2/f_\mathrm{s})$ is fed into a (hypothetical) digital-to-analogue converter that outputs a constant voltage $y(t) = y_n$ during the time interval $n/f_\mathrm{s} \leq t < (n+1)/f_\mathrm{s}$ for all integers n.

(*i*) Explain how this behaviour of the digital-to-analogue converter affects the amplitude spectrum of the resulting signal. [5 marks]

(*ii*) What amplitude A has to be chosen for the second term such that the resulting amplitude spectrum shows equally high peaks at both $f_1 = 1$ kHz and $f_2 = 2$ kHz if the sampling frequency is $f_\mathrm{s} = 6$ kHz? [5 marks]

11 Computer Systems Modelling

(*a*) Consider a general birth–death process with birth rate λ_i and death rate μ_i in state i ($i = 0, 1, 2, \ldots$). What are the detailed balance equations for this process? [2 marks]

(*b*) Derive the steady-state distribution for the general birth–death process considered in part (*a*). What are the conditions for the steady-state distribution to exist? [4 marks]

(*c*) Describe the M/M/1 queue and give a stochastic model for the number of customers present. Find the steady-state distribution for the number of customers present and state the condition for it to exist. [4 marks]

(*d*) Derive the mean number of customers present in steady state in the M/M/1 queue. [4 marks]

(*e*) State Little's law and use it to derive the mean time spent in the M/M/1 queue in the steady state. [2 marks]

(*f*) Discuss what is meant by the traffic intensity for an M/M/1 and explain what happens to the mean number of customers present as the traffic intensity increases towards one. [4 marks]

12 Quantum Computing

(*a*) Consider a quantum finite automaton with two basis states, $|0\rangle$ being the start state and $|1\rangle$ the only accepting state. The automaton operates on a two-letter alphabet, with matrices $M_a = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$ and $M_b = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

What are the probabilities that the automaton accepts each of the following input strings?

(*i*) a [3 marks]

(*ii*) aba [3 marks]

(*iii*) abb [3 marks]

(*b*) Give a complete description of the probabilities of acceptance associated with various possible input strings. [5 marks]

(*c*) Prove that there is no *two-state* probabilistic automaton with the same behaviour as the automaton described in part (*a*). [6 marks]

13 Bioinformatics

(*a*) Describe a bioinformatics application of hidden Markov models. [6 marks]

(*b*) Discuss the properties of the Markov clustering algorithm and the differences with respect to the k-means and hierarchical clustering algorithms. [8 marks]

(*c*) Describe the Gillespie algorithm and discuss its relationship with genetic or biochemical networks (give one example). [6 marks]

14 Types

(*a*) Give the rules inductively defining the type system of the *polymorphic lambda calculus* (PLC). [5 marks]

(*b*) What does it mean for a PLC expression M to be in *beta-normal form*? [2 marks]

(*c*) The *long normal forms*, L, and the *neutral forms*, N, are special kinds of PLC expression given by the following grammar:

$$\begin{aligned} L &::= \lambda x : \tau(L) \mid \Lambda\alpha(L) \mid N, \\ N &::= x \mid N\,L \mid N\,\tau. \end{aligned}$$

(*i*) Arguing by induction on the structure of L and N, or otherwise, show that all such expressions are in beta-normal form. [4 marks]

(*ii*) Show that if N is a neutral form, then $\{\} \vdash N : \tau$ is not provable in the PLC type system for any type τ, where $\{\}$ is the empty typing environment. (You may assume without proof that if $\Gamma \vdash M : \tau$ is provable in the PLC type system, then the free variables of the expression M are contained in the domain of definition of the typing environment Γ.) [3 marks]

(*iii*) Hence prove that for any long normal form L, $\{\} \vdash L : \forall\alpha(\alpha)$ is not provable in the PLC type system. [6 marks]

15 Denotational Semantics

Assume a denotational semantics is given for PCF: a mapping of PCF types τ to domains $[\![\tau]\!]$ together with a mapping of closed PCF terms of type τ, $M : \tau$, to elements of the corresponding domain $[\![M]\!] \in [\![\tau]\!]$. In the following you may refer to the notion of a PCF context without an explicit definition.

(*a*) State clearly what properties the semantics must fulfil in order to be compositional. Give the fundamental property that a compositional semantics enjoys with respect to arbitrary contexts. [2 marks]

(*b*) Assume the standard typed evaluation operational semantics for PCF. Without referring to explicit evaluation rules or to the syntax of PCF, state clearly what properties the denotational semantics must fulfil in order to be:

(*i*) sound; [1 mark]

(*ii*) adequate. [2 marks]

(*c*) Define:

(*i*) the contextual preorder $\leq_{\text{ctx}}$ for closed PCF terms; [2 marks]

(*ii*) contextual equivalence $\cong_{\text{ctx}}$ for closed PCF terms. [1 mark]

(*d*) Prove that if a denotational semantics is compositional, sound and adequate then denotational equality implies contextual equivalence. [4 marks]

(*e*) State the extensionality properties of the contextual preorder for both base types (*bool* and *nat*) and function types. [4 marks]

(*f*) Using extensionality, prove that the following terms are contextually equivalent:

(*i*) $(\mathbf{fn}\, x : \mathit{bool} \to \mathit{bool}\,.\, \mathbf{0})(\mathbf{fn}\, x : \mathit{bool}\,.\, \mathbf{true})$ and $\mathbf{0}$; [2 marks]

(*ii*) $\mathbf{fn}\, x : \mathit{nat}\,.\, \Big((\mathbf{fn}\, y : \mathit{nat} \to \mathit{nat}\,.\, y(yx))(\mathbf{fn}\, z : \mathit{nat}\,.\, \mathbf{succ}(z))\Big)$
and
$\mathbf{fn}\, x : \mathit{nat}\,.\, \mathbf{succ}(\mathbf{succ}(x))$. [2 marks]

16 Optimising Compilers

(*a*) Summarise very briefly (one short paragraph each) the overall idea behind program analysis using

(*i*) abstract interpretation;

(*ii*) set-constraint-based (CFA-like) analysis;

(*iii*) rule-based analysis.

[6 marks]

(*b*) Consider the following language of integer expressions e and (integer) list expressions E where n represents integer constants, x and X respectively range over integer and list variables, $\oplus$ represents integer operations (e.g. $+$, $\leq$ etc.), and **if** and **IF** test their first argument for zero/non-zero as in C:

$$\begin{aligned} e &::= x \mid n \mid e_1 \oplus e_2 \mid \mathbf{hd}\ E \mid \mathbf{if}(e_0, e_1, e_2) \\ E &::= X \mid [\,] \mid e :: E \mid \mathbf{tl}\ E \mid \mathbf{IF}(e, E_1, E_2) \end{aligned}$$

In *escape analysis and optimisation*, given a call to f such as

$$g(x, y) = f(x :: x :: [\,], y :: [\,])$$

we want to know whether or not the result of f can include any of the cons-cells reachable from its arguments. A formal parameter of f that might be incorporated into its result is known as *escaping*. This is useful because if (say) formal parameter 1 to f cannot escape then cons-cells allocated for actual parameter 1 can be allocated (more cheaply) on the stack instead of in the heap.

This problem may be formulated as an analysis that takes an expression, e or E, constituting the body of f. The parameters of f are the free variables, x_i and X_i, of its body.

Express this analysis using *two* of the techniques from part (*a*). In both cases state how to use the analysis result for e or E to test "parameter X_i definitely does not escape from E or e". [Hint: in some analyses it is easier to treat the variables x_i and X_i just as strings, and in others as variables ranging over $\{0, 1\}$.] [5 marks each]

(*c*) Indicate what changes would be necessary for *one* of your analyses were the syntax also to allow a recursive call to f. [4 marks]

END OF PAPER