COMPUTER SCIENCE TRIPOS Part II

Wednesday 6 June 2012 1.30 to 4.30

COMPUTER SCIENCE Paper 8

*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*) In at most five sentences, describe the *kd-tree* data structure and its use in ray tracing. [3 marks]

(*b*) Several formulae were discussed in the course for calculating the *normal to a discrete surface* at a vertex.

(*i*) In at most five sentences, explain why all of these formulae are necessarily approximations; and why a true answer is impossible. [3 marks]

(*ii*) Give the best (most nearly accurate) formula discussed in the course. [1 mark]

(*c*) The *angle deficit* is a method for measuring discrete curvature. Give the formula for the *angle deficit* of a vertex v. [1 mark]

(*d*) A simple closed discrete surface has 148 vertices, 248 edges and 80 faces. What must its genus be, and why? [2 marks]

(*e*) An implicit surface system is described by the force function $f(r) = 1/r^2$, the threshold $t = 1.0$, and a set of generating points.

(*i*) A surface is defined by two generating points at $(-x, 0, 0)$ and $(x, 0, 0)$. What is the largest value of x such that the surface forms a single component? [2 marks]

(*ii*) A surface is defined by three generating points, positioned in an equilateral triangle at $(-1, 0, 0)$, $(1, 0, 0)$, and $(0, \sqrt{3}, 0)$. Does the surface form a single component of genus 0, a single component of genus 1, or three separated components of genus 0? [1 mark]

(*iii*) If the set of generating points is described as $v \in \mathbf{V}$ we can write the equation for the total value of the force $F(x)$ at a point x in space as

$$F(x) = \sum_{v \in \mathbf{V}} \frac{1}{|x - v|^2}$$

Give an expression for $F(x)$ if the surface were instead defined by infinite *generating lines* of force, where each generating line, L, is expressed by two points $[A, B]$. If your expression is correct, the implicit surface described by a single line $[A, B]$ will be an infinite cylinder of radius one centred on the axis $[A, B]$. [3 marks]

(*iv*) Modify your definition to treat each generating line as a finite line segment from A to B. A single generating line should describe a surface that is a finite cylinder with spherical end caps. [2 marks]

(*v*) One alternative force function, for use with generating points, is:

$$f(r) = \begin{cases} 1 - \frac{4}{9}\frac{r^6}{R^6} + \frac{17}{9}\frac{r^4}{R^4} - \frac{22}{9}\frac{r^2}{R^2} & r < R \\ 0 & r \geq R \end{cases}$$

Suggest why and when this function is more computationally efficient than the function $f(r) = 1/r^2$. [2 marks]

2 Artificial Intelligence II

Evil Robot hates kittens. He has invented the *kitty-destroyer* (KD) to rid the world of their menace. To test it, he has a kitten (K) next to the KD on his laboratory bench (B). He has to open the KD, place the kitten inside it, close it and press the start button (SB). He has not however established that this sequence of events will lead to his goal of a destroyed kitten. Evil Robot is equipped with a planning system based on a solver for *constraint satisfaction problems*, and wants to use this to construct a plan.

(*a*) Explain how this problem can be represented using the *state-variable* representation, including in your answer specific examples of a *domain*, a *rigid relation*, a *state variable* and an *action* for the problem. [7 marks]

(*b*) Give *one* reason that a state-variable representation might be preferable to a representation aimed at encoding to a satisfiability problem. [1 mark]

(*c*) Explain, giving a specific example for this problem, how the *action* taken at some time t can be encoded as part of a constraint satisfaction problem. [3 marks]

(*d*) Explain, giving a specific example for this problem, how a *state-variable* can be encoded as part of a constraint satisfaction problem. [4 marks]

(*e*) Explain, giving a specific example for this problem, how a *precondition* for an action can be encoded as part of a constraint satisfaction problem. [5 marks]

3 Comparative Architectures

(*a*) Two independent threads are run on different processors in a chip-multiprocessor. Each thread simply increments a private counter one million times. The two counters are stored in consecutive memory locations. It is discovered that running the threads sequentially is faster than running them in parallel. What may cause this type of behaviour? [5 marks]

(*b*) Cache coherence protocols are classified as either invalidate or update protocols. What are the potential advantages and disadvantages of adopting an update rather than an invalidate protocol? [5 marks]

(*c*) Sequential consistency offers a simple and intuitive memory consistency model. Why is sequential consistency rarely supported by modern chip-multiprocessor designs? [5 marks]

(*d*) What information does the directory provide in a directory-based coherence protocol? [5 marks]

4 Computer Systems Modelling

(a) Let U be a uniform random variable on the interval $(0, 1)$. Show that for any continuous distribution function $F(x)$ the random variable X defined by

$$X = F^{-1}(U)$$

has the probability distribution function $F(x)$. [4 marks]

(b) Use your result in part (a) together with a random variable U distributed according to a uniform distribution on the interval $(0, 1)$ to construct random variables for the following two distributions:

(i) the uniform distribution on the interval (a, b) where a and b are real numbers such that $a < b$ [3 marks]

(ii) the exponential distribution with parameter $\lambda > 0$ [3 marks]

(c) Suppose that $X_1, X_2, \ldots, X_n$ are independent, identically distributed random variables with mean μ and variance σ^2. Use the central limit theorem to derive an approximate $100(1 - \alpha)$ percent confidence interval for μ. [5 marks]

(d) How would you obtain a confidence interval similar to that given in part (c) that is exact in the special case where the random variables $X_1, X_2, \ldots X_n$ have a Normal distribution? [5 marks]

5 Computer Vision

(*a*) The following very useful operator is often applied to an image $I(x, y)$ in computer vision algorithms, to generate a related "image" $g(x, y)$ for analysis:

$$g(x,y) = \int_\alpha \int_\beta \nabla^2 e^{-\left((x-\alpha)^2+(y-\beta)^2\right)/\sigma^2} I(\alpha,\beta)\ d\alpha\ d\beta$$

where

$$\nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$$

(*i*) Give the general name for this type of mathematical operation, and the chief purpose that it serves in computer vision. [1 mark]

(*ii*) What structures are expected at places (x, y) in the image $I(x, y)$ where the operator output $g(x, y)$ undergoes zero-crossings? [1 mark]

(*iii*) What is the significance of the parameter σ? If you increased its value, would there be more or fewer points (x, y) where $g(x, y) = 0$? [2 marks]

(*iv*) Describe the effect of the above operator in terms of the two-dimensional Fourier domain. What is the Fourier terminology for this image-domain operator? What are its general effects as a function of frequency, and as a function of orientation? [2 marks]

(*v*) If the computation of $g(x, y)$ were to be implemented entirely by Fourier methods, would its complexity be greater or less than the image-domain operation, and why? What would be the trade-offs involved? [2 marks]

(*vi*) If the image $I(x, y)$ has 2D Fourier Transform $F(u, v)$, provide an expression for $G(u, v)$, the 2D Fourier Transform of the operator output $g(x, y)$ in terms of $F(u, v)$, the Fourier plane variables u, v, some constants, and the parameter σ. [2 marks]

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

(*i*) "signal-to-symbol converter" [2 marks]

(*ii*) "inverse graphics" [2 marks]

(*iii*) quadrature demodulator [2 marks]

(*iv*) volumetric coordinates [2 marks]

(*v*) correspondence problem [2 marks]

6 Digital Signal Processing

BBC Radio Cambridgeshire broadcasts a radio signal on a carrier frequency of 1026 kHz with a (double-sided) bandwidth of 10 kHz. You connect a long wire (antenna) via an amplifier and bandpass filter (0.5–2.0 MHz) to an analog-to-digital converter (ADC) with a sampling frequency of 5 MHz.

(*a*) You record $n = 500$ consecutive samples $(x_0, x_1, \ldots, x_{499})$ from the analog-to-digital converter output and calculate the Discrete Fourier Transform (DFT)

$$X_k = \sum_{i=0}^{n-1} x_i \cdot \mathrm{e}^{-2\pi \mathrm{j} \frac{ik}{n}}$$

(*i*) For which index value(s) k do you expect $|X_k|$ to best indicate the received signal strength of this radio station? [4 marks]

(*ii*) What preprocessing step would improve this indication? [4 marks]

(*iii*) What redundancy do you expect to find in the DFT output vector X, considering that the input signal is real-valued? [2 marks]

(*iv*) Explain a technique that exploits this redundancy to calculate this real-valued DFT more efficiently. (You can assume that an FFT implementation is already available and that $n = 512$ is used instead.) [4 marks]

(*b*) You want to record the output of this radio station for later analysis, but you do not yet know how it was modulated. How can you convert the ADC output sequence $\{x_i\}$ such that the resulting sequence encodes efficiently what is happening in the frequency range 1021–1031 kHz, with as low a sample rate as possible? [4 marks]

(*c*) You finally learn that the signal recorded in (*b*) was an amplitude-modulated positive mono audio signal. How can you demodulate it? [2 marks]

7 E-Commerce

(*a*) Outline five pieces of information required by the Distance Selling Regulations. [5 marks]

(*b*) Explain what is meant by the "Freemium model" for online services. [5 marks]

(*c*) Explain the business model for social networks. [5 marks]

(*d*) Discuss whether newspapers can survive in a digital age. [5 marks]

8 Hoare Logic

(a) Describe two ways in which the interpretations of $\{P\}C\{Q\}$ differ in Hoare logic and in separation logic. [4 marks]

(b) Which of the following equations hold for all P and Q? In each case justify your answer.

(i) $P \star P = P$ [2 marks]

(ii) $P \wedge P = P$ [2 marks]

(iii) $P \star Q = Q \star P$ [2 marks]

(iv) $P \wedge Q = Q \wedge P$ [2 marks]

(c) Is the Hoare triple $\{\mathtt{X} \doteq 0 \star \mathtt{Y} \doteq 1\}$ `[X]:=Y` $\{\mathtt{X} \mapsto 1 \star \mathtt{Y} \doteq 1\}$ true? If so give a proof, stating any rules that you use; if not, explain why not. [4 marks]

(d) Is the Hoare triple $\{\mathtt{X} \mapsto 0 \star \mathtt{Y} \doteq 1\}$ `[X]:=Y` $\{\mathtt{X} \mapsto 1 \star \mathtt{Y} \doteq 1\}$ true? If so give a proof, stating any rules that you use; if not, explain why not. [4 marks]

9 Information Retrieval

The documents in the figure below are to be clustered according to their similarity in standard frequency-based vector space. The proximity metric to be used is the Manhattan distance $m(\vec{d_k}, \vec{d_j}) = \sum_i |d_{k,i} - d_{j,i}|$, where $\vec{d_k}$ and $\vec{d_j}$ represent the vectors assigned to documents k and j, and $d_{k,i}$ gives the frequency of term i in document k.

Doc 1: whale, sea, sea, whale, boat, boat, boat, boat, boat
Doc 2: whales, sea, sea, water
Doc 3: whale, water, water, whale, whale
Doc 4: whales, whales, whales

(a) Construct the term–document matrix under the assumption that the terms are not stemmed. [3 marks]

(b) Construct the corresponding document–document matrix. [3 marks]

(c) On the basis of the document–document matrix, perform complete-link clustering, showing the output as well as intermediate results. [6 marks]

(d) Starting from the situation in part (c), you now want to create a clustering which is guaranteed to be different from the one in (c). You are allowed to manipulate one of the following factors:

- the term weighting
- the proximity metric
- whether stemming is applied
- adding new terms to documents
- the similarity function (single-link instead of complete-link)

Which of the factors do you choose, and why? Demonstrate the changes affected. [8 marks]

10 Principles of Communications

(*a*) *Randomness* is used in several places in communication protocols. Discuss the role of random choices with regard to resource allocation, making reference to telephony call routing, Ethernet Media Access Control, and Queue Management in IP routers. [15 marks]

(*b*) Some researchers have claimed that RED (Random Early Discard) queueing can reduce the latency through a bottleneck router for all users.

(*i*) How might this be the case? [3 marks]

(*ii*) How could the addition of Explicit Congestion Notification help make it so? [2 marks]

11 Quantum Computing

(*a*) (*i*) Write out the matrix for the three-qubit Toffoli gate. [2 marks]

(*ii*) A three-qubit system is in the state

$$\frac{1}{2}(|000\rangle + |101\rangle + |110\rangle + |111\rangle).$$

The Toffoli gate is applied to this system and then a measurement is performed on the first qubit. What is the probability of observing a $|1\rangle$? Show all working out. [2 marks]

(*b*) State and prove the *no-cloning theorem*. [6 marks]

(*c*) Suppose we are given a qubit $|\theta\rangle$ in an unknown state, but we know it is either $|0\rangle$ or $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. We would like to devise a circuit to determine, with certainty, which of the two is the case, using any number of unitary and measurement operations.

Either explain how to construct such a circuit, or prove that no such circuit is possible. [10 marks]

12 Security II

The lifecycle of an exam question in a fictitious university includes at least the following stages, which take place over several months:

1. Professor invents question.
2. Chief examiner sanity-checks it.
3. Professor amends it if necessary.
4. External auditor sanity-checks it.
5. Professor amends it again if necessary.
6. Chief examiner approves final version.
7. Clerk prints question in required number of copies.

Following a scandal whereby some dishonest candidates got hold of questions ahead of time, thus forcing the whole exam to be invalidated and repeated to the dismay of the honest participants, the university has put pressure on its departments to ensure this will not happen again.

(a) The Head of Department A, where the leak occurred, is now paranoid about computer networks and insists that no exam question shall ever reside on any networked computer system until after the corresponding exam takes place.

(i) Describe four ways that a determined undergraduate might nonetheless get hold of exam questions before the exam even if that requirement were observed. [4 marks]

(ii) Describe a security policy suitable for department A, taking into account the head-of-department's requirements and the staff workflow. Discuss it thoroughly, including requirements analysis, incentives and technical mechanisms. [4 marks]

(b) The Head of Department B finds that A's requirement would impose an excessive penalty on the productivity of her staff. At the same time, she certainly does not want to be blamed for the next leak.

(i) Describe a security policy suitable for department B, taking into account the head-of-department's requirements and the staff workflow. Discuss it thoroughly, including requirements analysis, incentives and technical mechanisms. [6 marks]

(ii) Describe three trade-offs between security and usability that you considered in devising the policy in (b)(i) and justify the choices you made. [6 marks]

13 System-on-Chip Design

The inner loop of an application program, to be implemented within a SoC, performs a lot of bit-intensive operations on 64-bit numbers. Two of the operations are counting the number of ones and reversing the bits.

(*a*) What design considerations should be taken into account when deciding whether to speed up the implementation of these operations using custom hardware? [5 marks]

(*b*) Summarise each of the following three means of coupling general custom hardware to a processor sub-system within a SoC:

(*i*) implemented as a custom instruction [2 marks]

(*ii*) implemented in a peripheral controlled using programmed I/O [2 marks]

(*iii*) implemented in an autonomous bus master [2 marks]

(*c*) Discuss whether each of the three means of part (*b*) could be appropriate for the bit-intensive operations mentioned above. [4 marks]

(*d*) Technology predictions are that chip transistor count can continue to grow provided a lower percentage of the chip is in use at any one time. How can large parts of application programs be implemented in custom hardware and would this match technology trends? [5 marks]

[*Note:* In all sections, marks will be awarded for sensible argument even if assumptions or results are incorrect.]

14 Types

(*a*) Give the Mini-ML typing rules for variables, boolean values and conditional expressions, function abstraction and application, and for expressions of the form `let` $x = M_1$ `in` M_2. [6 marks]

(*b*) Midi-ML is obtained from Mini-ML by adding a unit type *unit*, reference types τ *ref* and associated expressions for the unit value (), reference creation `ref` M, dereferencing $!M$ and assignment $M_1 := M_2$. Give the Midi-ML typing rules for these forms of expression. [4 marks]

(*c*) What is meant by the *type soundness* property of a programming language and its type system? [1 mark]

(*d*) Explain why combining the typing rules in part (*a*) with those in part (*b*) leads to an unsound type system and how Midi-ML modifies the typing rule for let-expressions in order to ensure the type soundness property. Illustrate your answer using the expression

$$\texttt{let}\, r = \texttt{ref}\, \lambda x(x)\, \texttt{in}\, (\lambda y(!r()))(r := \lambda x(\texttt{if}\, x\, \texttt{then}\, \texttt{false}\, \texttt{else}\, \texttt{true})).$$

(You need not give a formal definition of the operational semantics of Midi-ML.) [9 marks]

END OF PAPER