COMPUTER SCIENCE TRIPOS Part II

Tuesday 4 June 2013 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
Graph paper
Blue cover sheets
Tags

SPECIAL REQUIREMENTS
Approved calculator permitted

1 Advanced Graphics

(*a*) A general piecewise curve definition, whether Bézier, B-spline, or NURBS, can be written as a sum of products of basis functions, $A_i(t)$, and control points, $\mathbf{P}_i$:

$$\mathbf{P}(t) = \sum A_i(t)\mathbf{P}_i,\ t_{\min} \leq t < t_{\max}$$

Give the conditions on the functions A_i that are needed to ensure that:

(*i*) Translation of all of the points by some vector, $\mathbf{P}'_i = \mathbf{P}_i + \Delta\mathbf{P}$, causes a translation of the curve by the same vector, $\mathbf{P}'(t) = \mathbf{P}(t) + \Delta\mathbf{P}$. [2 marks]

(*ii*) The curve lies within the convex hull of the control points. [2 marks]

(*iii*) The curve passes through one of the control points, $\mathbf{P}_j$. [2 marks]

(*b*) The knot vector $[0, 0, 0, 1, 1, 1]$ defines a quadratic B-spline with three control points. Derive the equations of and graph the three basis functions from this knot vector. [6 marks]

(*c*) The basis functions derived in part (*b*) can be used, in a NURBS curve, to reproduce exactly a quarter-circle. Recall that a NURBS curve can be written as:

$$\mathbf{P}(t) = \frac{\sum_{i=1}^{n+1} N_{i,k}(t)\mathbf{P}_i h_i}{\sum_{i=1}^{n+1} N_{i,k}(t) h_i},\ t_{\min} \leq t < t_{\max}$$

where h_i is the *homogeneous co-ordinate* associated with point $\mathbf{P}_i$. Place the three control points at $\mathbf{P}_1 = (1, 0)$, $\mathbf{P}_2 = (1, 1)$, $\mathbf{P}_3 = (0, 1)$.

(*i*) Sketch the NURBS curve for the case $h_1 = h_2 = h_3 = 1$ [1 mark]

(*ii*) Calculate the magnitude of the maximum error between the curve in (*c*)(*i*) and a perfect circle of radius 1 centred at $(0, 0)$. [2 marks]

(*iii*) Sketch the NURBS curve for the case $h_1 = h_3 = 1$, $h_2 = 0$. [1 mark]

(*iv*) Sketch the NURBS curve for the limit case $h_1 = h_3 = 1$, $h_2 \to \infty$. [1 mark]

(*v*) Derive the value for h_2 that makes the NURBS curve perfectly match a quarter circle of radius 1 centred at $(0, 0)$. [3 marks]

2 Artificial Intelligence II

Princess Precious is a very light sleeper, and insists that every night she must sleep on brand new silk sheets. Her younger brother, however, is in the habit of secretly scattering toast crumbs in her bed, to make sure she sleeps badly.

In order to get the week started well, he does this every Sunday with probability 0.9. For the rest of the week, he tends to relent on any given night if he placed crumbs in her bed the previous night, and hence leaves them with a probability of 0.1. On the other hand, if he did not leave crumbs on a given night, his mischievous nature compels him to leave crumbs the next night with probability 0.6.

Precious, being a true princess, tends to be grumpy in the morning if she has not slept well. Consequently, if she has slept with crumbs in her bed she is grumpy with probability 0.95. Being a light sleeper, even if there are no crumbs she is grumpy with probability 0.55.

(*a*) Give a detailed definition of a *Hidden Markov Model (HMM)* and show how the scenario described can be modelled as an HMM. [4 marks]

(*b*) Give a detailed description of the *Viterbi algorithm* for computing the most probable sequence of states, given that an HMM produces a given sequence of observations. [8 marks]

(*c*) It is observed that Princess Precious is grumpy on Monday and Tuesday. However on Wednesday she is radiantly happy. Use the Viterbi algorithm to compute the most likely sequence of activities performed by her brother. [8 marks]

3 Bioinformatics

Given the two DNA sequences: GCACTT and CCCAAT

(*a*) Compute the alignment (using the edit graph) and the final score with the following rules: match score = +1, mismatch = −1, gap penalty = −1. [4 marks]

(*b*) Discuss how the alignment score and the quality of the result depend on the match score, mismatch, and gap penalty. [6 marks]

(*c*) Generate four, short DNA sequences (a,b,c,d) such that their relations as a tree are approximately the following: ((a,b),(c,d)). [5 marks]

(*d*) How is the score matrix used in phylogenetic tree building techniques? [5 marks]

4 Business Studies

This question is about Intellectual Property and Copyright.

(*a*) List 5 types of Intellectual Property. Comment on their use to protect a software program. [5 marks]

(*b*) Explain what Intellectual Property Rights (IPR) an employee has on their inventions. Does this differ if the invention is made in their own time? [5 marks]

(*c*) Professor Elbowpatch, a professor of ancient languages and keen amateur gardener, invents a new sort of lawnmower. What are his IPR options? [5 marks]

(*d*) Discuss what actions Professor Elbowpatch should take to exploit his invention. [5 marks]

5 Comparative Architectures

(*a*) Why might a heterogeneous or asymmetric chip-multiprocessor be preferable to a homogeneous or symmetric one? [5 marks]

(*b*) You are a computer architect working on the design of a new processor for the mobile phone market. An initial analysis of applications suggests that there would be worthwhile gains in producing a processor that could offer two different power-performance tradeoffs. The first configuration would maximise the exploitation of ILP and consume the most power. The second would on average perform less well, but would consume less power.

(*i*) Describe how the microarchitecture of a single processor could be modified in order to offer the ability to switch between the two configurations described at run-time. [9 marks]

(*ii*) How might one determine when to switch from one configuration to the other in order to reduce overall power consumption while minimising the impact on the user experience? [6 marks]

6 Denotational Semantics

(a) Let D be a poset and let $f : D \to D$ be a monotone function.

(i) Give the definition of the least pre-fixed point, $\mathit{fix}(f)$, of f. Show that $\mathit{fix}(f)$ is a fixed point of f. [5 marks]

(ii) Show that whenever D is a domain and f is a continuous function, $\mathit{fix}(f)$ exists. [5 marks]

(b) A poset $(P, \sqsubseteq)$ has *binary meets* if for every pair of elements $x, y \in P$ there is a necessarily unique element $(x \sqcap y) \in P$ such that

- $(x \sqcap y) \sqsubseteq x$ and $(x \sqcap y) \sqsubseteq y$, and
- for all $z \in P$, $z \sqsubseteq x$ and $z \sqsubseteq y$ imply $z \sqsubseteq (x \sqcap y)$.

(i) Let $(P, \sqsubseteq)$ be a poset with binary meets. Show that the function $\mathit{meet} : P \times P \to P$ given by $\mathit{meet}(x, y) = x \sqcap y$ is monotone. [5 marks]

(ii) Exhibit a domain with binary meets for which the function *meet* is not continuous. Justify your answer. [5 marks]

7 Hoare Logic

(*a*) Briefly explain the concepts: mechanised program verification and verification conditions (VCs). [4 marks]

(*b*) Consider three consecutive assignments:

$$\{\ P\ \}\ \mathtt{V_1\ :=\ E_1;\ V_2\ :=\ E_2;\ V_3\ :=\ E_3}\ \{\ Q\ \}$$

Write down the VCs that are generated for such a program. Give a detailed proof which shows that, if the VCs are true, then the specification above is provable in Hoare Logic. [6 marks]

(*c*) Write down the VCs for the following annotated program. For this part, do *not* attempt to define *Inv*. [4 marks]

```
{ T }
I := 0;
X := 0;
Y := 1;
WHILE (I ≠ N) DO { Inv }
  I := I + 1;
  X := X + Y;
  Y := X + Y
OD
{ X = fib(2 × N) }
```

Here $\mathit{fib}(0) = 0$, $\mathit{fib}(1) = 1$ and $\mathit{fib}(n + 2) = \mathit{fib}(n) + \mathit{fib}(n + 1)$ for $n \in \mathbb{N}$.

(*d*) Provide a definition of *Inv* such that the VCs are provable. Sketch a proof of the VCs. [6 marks]

8 Human–Computer Interaction

Imagine you have been commissioned to design the user interface for a head-up display (e.g. based on Google Project Glass) that can be used while riding a bicycle, as a reminder of appointments around Cambridge.

(*a*) In order to be safe while riding, the visual design of appointment reminders and instructions should be as simple as possible. Describe three specific ways this can be achieved, using formal elements of visual design. [6 marks]

(*b*) Consider the possibility that users might wish to modify their appointment schedules while riding. Choose three different Cognitive Dimensions of Notations, and discuss their implications. [6 marks]

(*c*) Describe ways that features of the bicycle itself might form the basis for

(*i*) a tangible user interface; and

(*ii*) an augmented reality interface

to this system. For each of these, explain what sensor processing would be involved.

[4 marks]

(*d*) How might these two alternative interfaces be compared experimentally? Describe the structure of the experimental design and procedure for analysis of the results. [4 marks]

9 Information Theory and Coding

(a) Consider an alphabet of 5 symbols whose probabilities are as follows:

A	B	C	D	E
$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{2}$

One of these symbols has been selected at random and you need to discover which symbol it is by asking 'yes/no' questions that will be truthfully answered.

(i) What would be the most efficient sequence of such questions that you could ask in order to discover the selected symbol? [2 marks]

(ii) By what principle can you claim that each of your proposed questions in the sequence is maximally informative? [2 marks]

(iii) On average, how many such questions will need to be asked before the symbol is discovered? What is the entropy of the symbol set? [2 marks]

(iv) Construct a uniquely decodable prefix code for the symbols. Explain why it is uniquely decodable and why it has the prefix property. [2 marks]

(v) Relate the bits in the code words forming your prefix code to the 'yes/no' questions that you proposed in (i). [2 marks]

(b) Explain how the bits in an IrisCode are set by phase sequencing. Discuss how quantisation of the complex plane into phase quadrants sets each pair of bits; why it is beneficial for quadrant codes to form a Gray Code; how much entropy is thereby typically extracted from iris images; and why such bit sequences enable extremely efficient identity searches and matching. [5 marks]

(c) Consider a noisy analog communication channel of bandwidth $\omega = 1$ MHz, which is perturbed by additive white Gaussian noise whose total spectral power is $N_0\omega = 1$. Continuous signals are transmitted across such a channel, with average transmitted power $P = 1{,}000$. Give a numerical estimate for the *channel capacity*, in bits per second, of this noisy channel. Then, for a channel having the same bandwidth ω but whose signal-to-noise ratio $\frac{P}{N_0\omega}$ is four times better, repeat your numerical estimate of capacity in bits per second. [5 marks]

10 Natural Language Processing

Consider the following context-free grammar:

```
S -> NP VP        N -> dog        V -> sees
NP -> Det N       N ->cat         V -> hates
VP -> V           N -> mouse      V -> sneezes
VP -> V NP        Det -> the
```

(*a*) Which of the following sentences are recognised by this grammar, and why? [4 marks]

(*i*) the dog sneezes the cat

(*ii*) the mouse hates

(*iii*) the cat the mouse hates

(*iv*) the mouse hates the mouse

(*b*) Modify the grammar so that the following sentence is now accepted in addition:

the dog the cat the mouse sees hates sneezes

Your modification should express the linguistic phenomenon as efficiently and elegantly as possible. Justify your choice. [6 marks]

(*c*) The semantics of natural language expressions can be expressed in first order predicate logic (FOPL). For instance, "the dog sneezes" can be approximately expressed as

$$\exists x\, dog(x) \cap sneeze(x)$$

Following this pattern, express the semantics of the sentence in part (*b*) in FOPL. [4 marks]

(*d*) Consider the following sentence:

the mouse that sees the cat that hates the dog that sneezes

Contrast this construction to the one in part (*b*) in terms of semantics and syntax. How would you modify the original grammar in part (*a*) to account for this construction? [6 marks]

11 Optimising Compilers

(*a*) Give a semantic notion of a variable being live at a program point, explaining why this is problematic to calculate. Now give a simpler-to-calculate notion of liveness and explain how it relates to the semantic notion. Formulate *dataflow equations* whose solution(s) give the liveness at each program point. You need only consider liveness of simple non-address-taken variables. [4 marks]

(*b*) Suppose we have a basic block of p simple statements. Give a formula relating the liveness on entry to the block to those of its q neighbouring blocks in the control flow graph. This formula naturally uses $O(p)+O(q)$ operations – justify this statement. It is claimed that this formula can be re-arranged to require only $O(q)$ time to calculate by only using one '$\cup$' and one '$\setminus$' operator. Determine whether this is true. [*Hint:* you may wish to consider examples, and to start by solving the case $p = 2$. Partial credit will be given for a good set of concrete examples arguing for or against.] [5 marks]

(*c*) To solve the dataflow equations, an initial approximation to liveness at the start of each basic block is required. What is it, and indicate why this leads to a preferable solution. [2 marks]

(*d*) Solving dataflow equations is usually expressed iteratively, where each iteration is of the form "for every basic block re-calculate the set of live variables from the current sets of live variables of its neighbours". We want to determine whether some basic-block orderings in "for every basic block" result in fewer overall iterations than others. Suppose the program has k basic blocks, but no cycle in the control flow graph; give an optimal ordering which only requires one dataflow iteration to calculate liveness (a second would only calculate the same value of the first). Also give such a program and an ordering which maximises the number of iterations required, giving the number of iterations in terms of k. [5 marks]

(*e*) Consider the program with four labelled blocks (with `B1` as entry node):

```
B1: x = read(); y = read(); z = read(); goto B2;
B2: z = z+1; x = x-1; if (x>0) goto B3; else goto B4;
B3: z = z+1; y = y-1; if (y>0) goto B2; else goto B4;
B4: print(z);
```

Show (*i*) there is no basic block ordering for which a single iteration gives the correct liveness at each label, but (*ii*) there is an ordering for which two iterations suffice (in the sense that a third would agree with the second). Give your ordering both explicitly as a permutation of $\{\texttt{B1}, \texttt{B2}, \texttt{B3}, \texttt{B4}\}$ and also as a general principle along the lines of your answer to part (*d*). [4 marks]

12 Principles of Communications

Suppose that you read about the design of an end-to-end transport protocol for an early version of the Internet, which uses window-based flow control with a fixed size window, with *Go-Back-N* retransmission of all un-acknowledged packets when there is a time-out awaiting an acknowledgement for any given data packet.

How would you convince the designer that they are going to have real problems with such a simplistic scheme? [20 marks]

13 Security II

(*a*) Formally state the two rules of the Bell-LaPadula (BLP) security policy model and then re-state them informally in terms of a single rule about the direction of information flow. [2 marks]

(*b*) Consider a distributed system in which A is a *TOP SECRET* process running on machine Alice and B is a *CONFIDENTIAL* object residing on machine Bob.

(*i*) Explain and justify whether A is allowed to read and/or write from B according to the BLP policy. [2 marks]

(*ii*) Discuss the claim made by some researchers that this scenario highlights a fundamental problem with the BLP policy. [4 marks]

(*c*) Consider the following description of Brewer and Nash's Chinese Wall security policy model.

- *Simple rule*: Read or write access to object o_2 by subject s is granted if and only if, for all objects o_1 to which s has had access, we have: (class(company(o_1)) $\neq$ class(company(o_2)) **or** (company(o_1) $=$ company(o_2)).

- **-rule*: Write access to object o_2 by subject s is granted if and only if access is granted by the *simple rule* and there does not exist any unsanitized object o_1, readable by s, for which company(o_1) $\neq$ company(o_2).

(*i*) Explain the context and goal of the Chinese Wall security policy model. Then explain what each of the two rules is intended to enforce or prevent. [4 marks]

(*ii*) Some researchers have claimed that the formal rules of Chinese Wall do not match the policy that Brewer and Nash intended to enforce, to the extent that the resulting policy is unusable in practice. Explain precisely why the policy would be unusable and give a clear proof of this claim. [8 marks]

14 Topical Issues

Consider an inertial measurement unit that reports its change in distance and heading each second, but is subject to typical sensor errors such as bias and noise. A particle filter is to be used to fuse its output with a building map to estimate the current location of a user within that building.

(*a*) Describe the evolution of the spatial distribution of the particles as a user moves through the building to their office. Assume that the initial position is unknown, but that the filter has correctly estimated the user's position before they reach their office. Additionally, the filter continues to run thereafter (during which time they are seated and not travelling). What factors influence the speed of transition between the regimes identified? [9 marks]

(*b*) The particle filter is to run on a low power embedded platform that transmits its position results over WiFi. However, it is found to process particles too slowly for real-time tracking.

(*i*) Describe the three stages of a particle filter cycle in the context described above. Assuming a multi-core GPU is available, discuss how easily they can be parallelised to increase performance. [5 marks]

(*ii*) Describe two further ways to increase performance in such a system. [4 marks]

(*c*) The internal structures of many office buildings change over time, as offices merge or split. Assuming the building map is not kept synchronised with these changes, discuss the effects on the positioning system described. [2 marks]

END OF PAPER