COMPUTER SCIENCE TRIPOS Part II

Tuesday 3 June 2014 1.30 to 4.30 pm

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
Rough work pad

SPECIAL REQUIREMENTS
Approved calculator permitted

1 Advanced Graphics

(*a*) Assuming that it is mapped to a square with texture co-ordinates from $(0, 0)$ to $(1, 1)$, sketch a picture of the procedural texture map generated by the following Java code. Use textual annotations to indicate the colours of the various parts of the picture. [7 marks]

```
Color BLUE = ...;  Color GREEN = ...;  Color WHITE = ...;
Color parametricTexture(double tu, double tv) {
  double r1 = 2 * Math.sqrt(  (tu - 0.25) * (tu - 0.25)
                            + (tv - 0.5)  * (tv - 0.5) );
  double r2 = 2 * Math.sqrt(  (tu - 0.75) * (tu - 0.75)
                            + (tv - 0.5)  * (tv - 0.5) );
  double f = f(r1) + f(r2);
  return (f > 0.45 && f < 0.55) ? WHITE :
          GREEN.times(f).plus(BLUE.times(1 - f));
}
private double f(double r) {
  if (0 <= r && r < 0.333)      { return 1 - 3*r*r; }
  else if (0.333 <= r && r < 1) { return 1.5*(1-r)*(1-r); }
  else                          { return 0; }
}
```

(*b*) Write brief notes that explain the following.

(*i*) Explain *Perlin Noise*, including how it differs from white noise. [3 marks]

(*ii*) Explain *Barycentric Co-ordinates*, including how they are calculated. Diagrams are encouraged. [3 marks]

(*c*) Given a ray $R(t) = O + Dt$ and a unit sphere S, initially centred on the origin and subsequently transformed by affine matrix M, where M represents the transformation of the sphere from local to world coordinates:

(*i*) state the centre of the sphere in local co-ordinates and in world co-ordinates;

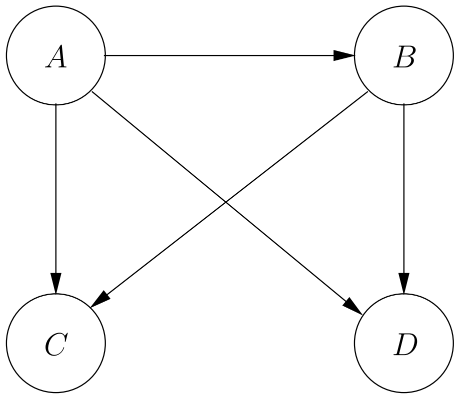
(*ii*) give an expression in terms of t for the local co-ordinates of the intersections between R and S;

(*iii*) give an expression for the world co-ordinates of the same intersections; and

(*iv*) give an expression for the world co-ordinates of the normal at those intersections. [7 marks]

2 Artificial Intelligence II

Consider the following Bayesian network:



The associated probability distributions for the binary random variables A, B, C and D are $\Pr(a) = 0.2$, $\Pr(\neg a) = 0.8$ and:

A	$\Pr(b\|A)$	A	B	$\Pr(c\|A,B)$	$\Pr(d\|A,B)$
$\bot$	0.8	$\bot$	$\bot$	0.4	0.9
$\top$	0.7	$\bot$	$\top$	0.2	0.8
		$\top$	$\bot$	0.3	0.1
		$\top$	$\top$	0.1	0.2

(a) Write down an expression for the full joint distribution of the random variables A, B, C and D. Compute the probability that A, B and C are $\bot$ while D is $\top$. [2 marks]

(b) Use the *variable elimination algorithm* to compute the probability distribution of C conditional on the evidence that $D = \bot$. [16 marks]

(c) Comment on the computational complexity of the variable elimination algorithm. [2 marks]

3 Bioinformatics

(*a*) Describe the differences in complexity and usage between parsimony and distance phylogenetic methods. Give an example of the usage of both methods. [6 marks]

(*b*) Describe the differences in complexity and usage between hierarchical clustering and the Markov clustering (MCL) algorithm. [5 marks]

(*c*) Explain how to identify different gene features using Hidden Markov Model methods such as Genescan. [5 marks]

(*d*) Explain how you could identify a regulatory network involving a set of genes. [4 marks]

4 Business Studies

(*a*) List, with one sentence descriptions, five methods of valuing a company. [5 marks]

(*b*) The outline profit and loss account and balance sheet for Certain Software Ltd are as follows:

£'000	Q1	Q2	Q3	Q4	TOTAL
INCOME					
Consultancy	100	200	300	400	1000
Licences	0	25	50	25	100
Total income	100	225	350	425	1100
EXPENDITURE					
Salaries	100	100	100	100	400
Overheads	125	125	125	125	500
Total expenditure	225	225	225	225	900
Profit/(loss)	(125)	0	125	200	200

Balance sheet

SOURCE	
Shareholders capital	100
Bank Loan	50
Trade Debtors	200
Retained Profit	200
Total	550
APPLICATION	
Assets: Computers, desks etc	200
Creditors	100
Bank	250
Total	550

(*i*) Provide three different valuations for the company, showing how they were obtained. State any assumptions that you make. [6 marks]

(*ii*) 75% of sales are made to Mammoth Corporation, a large public US company. The company receives an offer from Mammoth of $1.5M for the company. Certain Software Ltd is owned by the founding Directors who still work for the company. Advise the directors. [5 marks]

(*iii*) As negotiations proceed it emerges that the takeover terms are $500K in cash and $1M in shares of Mammoth Corporation, locked-in for 1 year. Does this alter your advice? [4 marks]

5 Comparative Architectures

(*a*) A 4KB, blocking, private L1 cache with 16B lines sees the following sequence of accesses from its core.

```
0x00001000 Load
0x00001010 Store
0x00002000 Load
0x00001010 Load
0x00003000 Load
0x00001010 Store
0x00001010 Store
0x00002000 Load
0x00001000 Load
0x00002000 Load
```

Assuming a write-allocate cache that is empty at first and implements the least-recently-used (LRU) replacement algorithm, what is the hit rate if the cache is

(*i*) direct-mapped;

(*ii*) fully-associative;

(*iii*) 2-way set-associative?

[6 marks]

(*b*) If the core supports out-of-order execution, how might a non-blocking cache bring performance benefits? [4 marks]

(*c*) How might the core's load/store queue be used to reduce the number of memory accesses seen by the cache? [4 marks]

(*d*) Assume that this core and cache are part of a chip-multiprocessor, with the cache connected to a shared L2 via a bus that maintains coherence through a snooping MESI protocol. What sequence of steps would be taken if another core wanted to load from 0x00001010 after the given sequence had finished?

[6 marks]

6 Denotational Semantics

For partially ordered sets $(P, \sqsubseteq_P)$ and $(Q, \sqsubseteq_Q)$, define the set

$$(P \Rightarrow Q) = \{f \mid f \text{ is a monotone function from } (P, \sqsubseteq_P) \text{ to } (Q, \sqsubseteq_Q)\}$$

and, for all $f, g \in (P \Rightarrow Q)$, let

$$f \sqsubseteq_{(P \Rightarrow Q)} g \iff \forall\, p \in P.\, f(p) \sqsubseteq_Q g(p)$$

(*a*) Let $(P, \sqsubseteq_P)$ and $(Q, \sqsubseteq_Q)$ be partially ordered sets.

(*i*) Prove that $\big((P \Rightarrow Q), \sqsubseteq_{(P \Rightarrow Q)}\big)$ is a partially ordered set. [4 marks]

(*ii*) Prove that if $(Q, \sqsubseteq_Q)$ is a domain then so is $\big((P \Rightarrow Q), \sqsubseteq_{(P \Rightarrow Q)}\big)$. [6 marks]

(*b*) For $\mathbb{N}$ the set of natural numbers partially ordered by the equality relation and for $S_\bot$ the flat domain determined by a set S, consider the domain $\big((\mathbb{N} \Rightarrow S_\bot), \sqsubseteq_{(\mathbb{N} \Rightarrow S_\bot)}\big)$.

(*i*) A function $f \in (\mathbb{N} \Rightarrow S_\bot)$ is said to be *finite* whenever the subset of $\mathbb{N}$ given by $\{n \mid f(n) \neq \bot\}$ is finite.

Show that every function in $(\mathbb{N} \Rightarrow S_\bot)$ is the least upper bound of a countable chain of finite functions. [4 marks]

(*ii*) For a domain $(D, \sqsubseteq)$, an element $d \in D$ is said to be *isolated* (with respect to $\sqsubseteq$) whenever, for all countable chains $(x_0 \sqsubseteq \cdots \sqsubseteq x_n \sqsubseteq \cdots)$ in D with $d \sqsubseteq \bigsqcup_{n \geq 0} x_n$, there exists $m \geq 0$ with $d \sqsubseteq x_m$.

Prove that a function in $(\mathbb{N} \Rightarrow S_\bot)$ is isolated (with respect to $\sqsubseteq_{(\mathbb{N} \Rightarrow S_\bot)}$) iff it is finite. [6 marks]

7 Hoare Logic

(*a*) Consider Hoare triples of the form $\{\mathtt{T}\}\ V\mathtt{:=}E\ \{V = E\}$ where $\mathtt{T}$ is the atomic formula 'true' and V and E range over variables and expressions, respectively.

(*i*) Write down an instance of such a triple that cannot be proved using Hoare logic and explain why not. [2 marks]

(*ii*) Write down conditions on V and E such that $\{\mathtt{T}\}\ V\mathtt{:=}E\ \{V = E\}$ can be proved and give a proof of this assuming your conditions. [2 marks]

(*b*) Write down and explain the weakest liberal precondition $\mathtt{wlp}(V\mathtt{:=}E,\ Q)$ and strongest postcondition $\mathtt{sp}(V\mathtt{:=}E,\ P)$. Comment on the relationship of these to the Hoare triple $\{P\}\ V\mathtt{:=}E\ \{Q\}$. [4 marks]

(*c*) Explain briefly how both weakest preconditions and strongest postconditions are used in mechanised program verification. [4 marks]

(*d*) Write down the Hoare assignment axiom and the Floyd assignment axiom. Explain carefully why each is true. [4 marks]

(*e*) Show how the Floyd assignment axiom can be derived from the Hoare assignment axiom and the other standard rules of Hoare logic. [4 marks]

8 Human–Computer Interaction

This question is concerned with the following design problem: Design a visualisation that will be useful to candidates sitting an exam such as the one you are sitting now. This visualisation should help the user to plan his or her time, allocating time to various questions. Elements of the visualisation might include estimated time to be spent on each question and part of a question; the amount of credit to be awarded for each part; the estimated likelihood of a correct answer; dependencies between parts (where correctness of one part depends on correctness of another); the remaining time; degree of fatigue; any other relevant factors, or any relevant subset of these. This visualisation could be used as the basis for an interactive screen-based assistant, but could also be incorporated in an augmented reality headset, or even implemented with pencil and paper.

(*a*) Make sketches of two alternative visual representations that might address this problem. If your design is intended to include coloured elements, colours may be indicated by text annotations on the sketch or with a separate key, rather than using coloured ink. The two alternative sketches should demonstrate different approaches to the problem. [4 marks]

(*b*) Contrast the design decisions in your two alternative sketches, discussing how the marks and regions on the page correspond to their intended meaning. [6 marks]

(*c*) Define what is meant by "analytic usability evaluation" and "empirical usability evaluation" techniques. [2 marks]

(*d*) Using an analytic evaluation technique, compare two different ways in which user experiences with your proposed designs would be expected to differ from each other. [4 marks]

(*e*) Describe how you would conduct an empirical evaluation to compare user experiences with your proposed designs. [4 marks]

9 Information Theory and Coding

(*a*) Suppose that women who live beyond the age of 80 outnumber men in the same age group by three to one. How much information, in bits, is gained by learning that a person who lives beyond 80 is male? [2 marks]

(*b*) Consider n discrete random variables, named $X_1, X_2, \ldots, X_n$, of which X_i has entropy $H(X_i)$, the largest being $H(X_L)$. What is the upper bound on the joint entropy $H(X_1, X_2, \ldots, X_n)$ of all these random variables, and under what condition will this upper bound be reached? What is the lower bound on the joint entropy $H(X_1, X_2, \ldots, X_n)$? [3 marks]

(*c*) If discrete symbols from an alphabet S having entropy $H(S)$ are encoded into blocks of length n symbols, we derive a new alphabet of symbol blocks S^n. If the occurrence of symbols is independent, then what is the entropy $H(S^n)$ of this new alphabet of symbol blocks? [2 marks]

(*d*) Consider an asymmetric communication channel whose input source is the binary alphabet $X = \{0, 1\}$ with probabilities $\{0.5, 0.5\}$ and whose outputs Y are also this binary alphabet $\{0, 1\}$, but with asymmetric error probabilities. Thus an input 0 is flipped with probability α, but an input 1 is flipped with probability β, giving this channel matrix $p(y_k|x_j)$:

$$\begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

(*i*) Give the probabilities of both outputs, $p(Y = 0)$ and $p(Y = 1)$. [2 marks]

(*ii*) Give all the values of (α, β) that would maximise the capacity of this channel, and state what that capacity then would be. [3 marks]

(*iii*) Give all the values of (α, β) that would minimise the capacity of this channel, and state what that capacity then would be. [3 marks]

(*e*) In order for a variable length code having N codewords with bit lengths $\{n_1, n_2, n_3, \cdots, n_N\}$ to satisfy the prefix property, what condition must be satisfied? (Express the condition, but do not try to prove it.) [1 mark]

(*f*) The information in continuous signals which are strictly bandlimited (lowpass or bandpass) is *quantised*, in that such continuous signals can be completely represented by a finite set of discrete samples. Describe two theorems about how discrete samples suffice for exact reconstruction of continuous bandlimited signals, even at all the points between the sampled values. [4 marks]

10 Natural Language Processing

The following context-free grammar (CFG) accepts sequences of part-of-speech categories (e.g., Det N, Adj Adj N). With a lexicon, as shown, it can be used to parse some English noun phrases (NPs).

```
Start symbol: NP        a, the: Det
NP -> Det N             dog, dogs, house,
NP -> N                   houses, model, models: N
N -> Adj N              brown, red, model: Adj
N -> N PP               in, under: P
PP -> P NP
```

(*a*) Give a non-deterministic finite-state automaton (NDFSA) which accepts the same sequences of part-of-speech categories as this CFG. Explain the notation that you use. [6 marks]

(*b*) Give two examples of overgeneration that can be demonstrated with the lexicon shown, and explain how the CFG and NDFSA (and, if necessary, part-of-speech categories and lexicon) could be modified to prevent them. [5 marks]

(*c*) The CFG does not accept noun-noun compounds (e.g., *the dog house*, *house dogs*). Indicate how you could modify the original CFG and NDFSA to allow for them. [3 marks]

(*d*) Hand-constructed FSA have sometimes been used for part-of-speech tagging. Outline the possible practical and theoretical advantages and disadvantages of such an approach when compared to stochastic tagging using Hidden Markov Models. [6 marks]

11 Optimising Compilers

(*a*) Explain the ideas behind *available expression analysis*. Your explanation should include data-flow equations, an informal argument as to why these correctly capture a semantic notion of availability, any issues with decidability and an algorithm to solve the data-flow equations. It is sufficient to consider, as candidate available expressions, those expressions of the form $v \oplus v'$ where $\oplus$ is a binary operation and v and v' are variables or constants. [6 marks]

(*b*) Show how the result of available expression analysis can be used to perform *common sub-expression elimination*. You need not give an algorithm, but explain the steps in the optimisation carefully. [4 marks]

(*c*) Assume that your intermediate code is represented by three-address instructions stored within basic blocks, and with a fresh temporary used whenever a temporary variable is used to hold intermediate results of a larger expression. Explain how your algorithm deals with optimising the program fragment

```
u = f(a+b*c, a+b*c);
v[a+b*c] = u;
```

where `a` is a global variable which may be updated by `f`, and `b` and `c` are local variables. [5 marks]

(*d*) Explain carefully how your common sub-expression elimination algorithm reacts to program fragment:

```
do { x += b*c; ... } while (...);
```

and also to program fragment:

```
z = b*c; do { x += b*c; ... } while (...);
```

commenting on any differences and on any similarity to lifting a loop-invariant expression out from a loop. In both cases assume neither `b` nor `c` is modified anywhere in the loop. [5 marks]

12 Principles of Communications

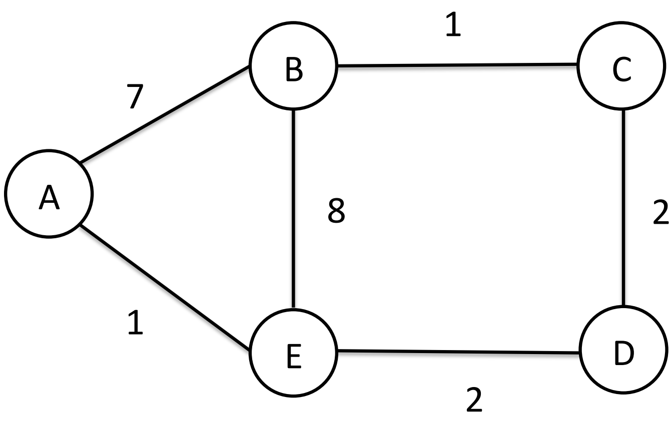
(*a*) The network illustrated below is running a distance vector routing algorithm, with update periodicity configured to 30 second intervals. The links are all 100Mbps Ethernets, including the host attachments to routers. The edge weights on the network are the (fixed) delays in milliseconds.

A source host connected to router A opens a TCP flow to a destination host connected to router B, and starts to move a file across the path, which has packet sizes (maximum segment size) of 1500 bytes, and an underlying RTT given by the sum of link metrics on a path (e.g. 2x6ms). Assume that prior to the start of the TCP flow, the network topology and routing have reached a steady state.

Suppose that the link between B and C fails 30 seconds after the TCP flow has started and the link is repaired after 5 minutes.

(*i*) Explain the route computations that follow the failure and the repair of the link. [8 marks]

(*ii*) Draw a time sequence diagram of TCP data and acknowledgement packets, illustrating the congestion window alongside. [8 marks]



(*b*) How much simpler would the computations have been if we ran Link State rather than Distance Vector routing? [4 marks]

13 Security II

(*a*) Consider a gamble in which you have even chances of winning or losing \$100. Your expected gain or loss will be \$0 and you should therefore be indifferent to gambling or not gambling. Explain Bernoulli's idea, at the foundation of Expected Utility Theory, that instead says you will be risk-averse. [4 marks]

(*b*) Kahneman's 2002 Nobel Prize Lecture features the following two problems:

Problem 2
Would you accept this gamble?
50% chance to win \$150
50% chance to lose \$100
Would your choice change if your overall wealth were lower by \$100?

Problem 3
Which would you choose?
lose \$100 with certainty
or
50% chance to win \$50
50% chance to lose \$200
Would your choice change if your overall wealth were higher by \$100?

Explain all the important points made by these problems, including why they are related. Show how they demonstrate "Bernoulli's error", in the context of Prospect Theory being a critique of Expected Utility Theory. [8 marks]

(*c*) Sketch the main graph of Prospect Theory and describe its most significant features, with reference to the two problems in (*b*). [4 marks]

(*d*) Explain what is meant by "tragedy of the commons" and how the concept relates to password usability. [2 marks]

(*e*) Explain what is meant by "compliance budget" and how the concept relates to password usability. [2 marks]

END OF PAPER