COMPUTER SCIENCE TRIPOS Part II

Wednesday 8 June 2005 1.30 to 4.30

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 Coversheets
Tags

1 Comparative Architectures

(*a*) All modern processors provide support for memory access protection and translation. Describe the Translation Lookaside Buffer (TLB) architecture of a modern microprocessor, and hence what information a typical TLB entry would contain. [6 marks]

(*b*) Some architectures are described as having software-managed TLBs, whereas others have entries loaded entirely by hardware. Describe and contrast the two approaches. [6 marks]

(*c*) Why might a TLB that supports "superpages" (entries that cover multiple pages) benefit applications that use lots of memory? What steps must the operating system take to enable superpages to be used? [4 marks]

(*d*) Even hardware-filled TLBs usually rely on software to determine when entries should be invalidated. How might you design a hardware solution to automatically keep the TLB coherent with OS pagetables? What extra complications would Symmetric Multiprocessor Systems (SMP) pose? [4 marks]

2 Artificial Intelligence II

(*a*) A given probabilistic inference problem involves a query random variable (RV) Q, evidence RVs $\mathbf{E} = (E_1, \ldots, E_n)$ and unobserved RVs $\mathbf{U} = (U_1, \ldots, U_m)$. Assuming that RVs are discrete, state the equation allowing the inference $\Pr(Q|\mathbf{E} = (e_1, \ldots, e_n))$ to be computed using the full joint distribution of the RVs and explain why in practice such a method might fail. [5 marks]

(*b*) Give a general definition of a *Bayesian network* (BN), and explain how a BN represents a joint probability distribution. [4 marks]

(*c*) Define *conditional independence* and explain how BNs make use of this concept to reduce the effect of the difficulties mentioned in your answer to part (*a*). Describe the way in which conditional independence is employed by the *naïve Bayes* algorithm. [6 marks]

(*d*) Describe *two* further issues relevant to the application of BNs in a practical context and describe briefly how these issues can be addressed. [5 marks]

3 Digital Communication II

(*a*) Outline the mechanism that most TCP implementations use today to set the retransmission timer dynamically. [10 marks]

(*b*) TCP uses additive increase multiplicative decrease (AIMD) for congestion avoidance in the steady state. Describe *two* optimisations that modern TCP variants use to improve performance when there is moderate packet loss. [5 marks]

(*c*) What might you propose to solve the problem of repeated slow-start of TCP in networks where packet loss due to noise was high (e.g. 50% packet loss probability), but there was still the possibility of congestion? [5 marks]

4 Distributed Systems

An appropriate structure for large-scale distributed systems is as multiple, independently administered, firewall-protected, domains. Examples are a national health service, a national police service and a global company with worldwide branches. Communication must be supported within and between domains and external services may be accessed. For example, health service domains may all access a national Electronic Health Record service; police service domains may all access a national Vehicle Licensing service.

(*a*) (*i*) Define publish/subscribe communication. [3 marks]

(*ii*) What are the advantages and disadvantages of offering publish/subscribe as the only communication service? [7 marks]

(*b*) (*i*) Define rôle-based access control. [3 marks]

(*ii*) What are the advantages and disadvantages of using rôle names for access control and communication? [7 marks]

Illustrate your discussions by means of examples.

5 Advanced Systems Topics

A computer system provides a compare-and-swap (CAS) operation which is used in the following manner:

```
seen = CAS (address, old, new)
```

It loads the contents of `address`, compares that value against `old` and if it matches stores the value `new` at the same address. All of this is performed atomically and the value read from the address is returned as `seen`.

(*a*) Making use of CAS, write pseudo-code for a simple multi-reader spin-lock. Your design should permit concurrent readers to enter their critical sections in parallel but ensure that writers gain exclusive access. Be sure to provide pseudo-code for each of the four operations supported by the lock, and describe the layout of the lock's data fields in memory. [10 marks]

(*b*) Why might this simple spin-lock perform poorly on a large multi-processor system? How might you improve the lock to achieve better performance on such a system? [4 marks]

A programmer analyses a multi-threaded application and discovers that the majority of the execution time is spent contending for access to a shared data structure.

(*c*) Describe *three* methods for reducing lock contention amongst threads accessing a highly-concurrent data structure. In each case briefly describe a situation or workload to which the method is particularly well suited. [6 marks]

6 Security

A mobile telephone company has 5 million prepay customers who buy scratchcards to pay for air time. Each card has a code of 9 decimal digits, and at any time there are about 20 million cards active (issued to the supply chain and not yet used).

(*a*) Discuss the relative advantages and disadvantages of implementing the code system with a database of random numbers *versus* an encrypted counter. [4 marks]

(*b*) If you were using an encrypted-counter system, how would you go about selecting, adapting or designing a suitable cipher? [4 marks]

(*c*) Some of the customers have got clever. As they are allowed two invalid code attempts, they try two random codes before entering a correct one. The telephone company is now getting 2000 complaints a month from people who bought a scratchcard and found, when they tried to use it, that someone else had already guessed the number. How would you modify the system to reduce the level of complaints? [8 marks]

(*d*) You are now approached by a telephone company in China which wants to use your system to manage 100 million prepay customers. What further modifications would you consider? [4 marks]

7 Optimising Compilers

(a) A dataflow analyser is required which can report on local variables having *write–write dataflow anomalies*. A write–write anomaly is present in a program if there is a path in the flowgraph containing two writes to a given variable and with no intervening read to that variable. For example

```
y=a; if (p) x=1; if (q) x=2; if (y==b) y=1; else y=2;
```

has an anomaly for `x` but not for `y`.

Given node n in the flowgraph, let $R(n)$ be the set of variables v for which a node n' exists with n' writing to v and having a path from n' to n which does not contain a read from v.

(i) Give dataflow equations for $R(n)$ and thence construct an algorithm which reports variables having such anomalies. Pay attention to the initialisation of any iteration which you employ. [8 marks]

(ii) Discuss briefly to what extent your algorithm could be extended to deal with global variables or with address-taken local variables. [4 marks]

(b) Let us say that an undirected graph (N, E) is k-cyclic if $N = \{n_1, \ldots, n_k\}$ and $E = \{(n_1, n_2), (n_2, n_3), \ldots (n_{k-1}, n_k), (n_k, n_1)\}$.

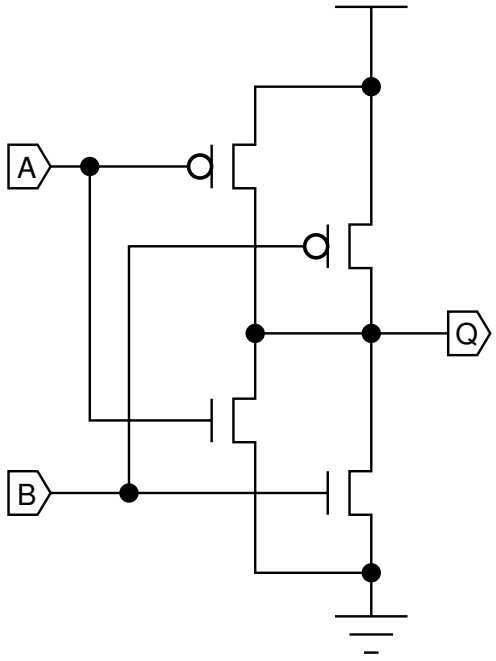
(i) Give a function body, or flowgraph, for which the register inference graph for its local variables forms a 4-cyclic graph. [4 marks]

(ii) Give a formula $c(k)$ which gives the number of colours (or registers) needed for a minimal colouring of a k-cyclic graph. [4 marks]

8 VLSI Design

(*a*) Sketch the circuit of a 2-input NOR gate in static CMOS, and explain how it works. [4 marks]

(*b*) Johnson's alternative design has the following circuit:



Explain how it works. [4 marks]

(*c*) Correct operation relies on careful choice of the transistor sizes. Assuming that a conducting p-channel has a resistance γ times that of a similarly sized n-channel and that the threshold voltage is $\frac{1}{5}$ the operating voltage, calculate suitable widths for the transistors, and explain the reasons for their values. [2 marks]

(*d*) Calculate the logical effort and parasitic delay for both designs when $\gamma = 2$. [10 marks]

9 Bioinformatics

(a) Present the aim of Blast software:

(i) Describe in words how the algorithm works. [5 marks]

(ii) Describe the output of a Blast search. [5 marks]

(b) Describe the aim of microarray data analysis:

(i) Describe the format of microarray data. [2 marks]

(ii) Describe in words how a cluster algorithm works. [8 marks]

10 Information Theory and Coding

(*a*) For continuous random variables X and Y, taking on continuous values x and y respectively with probability densities $p(x)$ and $p(y)$ and with joint probability distribution $p(x, y)$ and conditional probability distribution $p(x|y)$, define:

(*i*) the *differential entropy* $h(X)$ of random variable X; [1 mark]

(*ii*) the *joint entropy* $h(X, Y)$ of the random variables X and Y; [1 mark]

(*iii*) the *conditional entropy* $h(X|Y)$ of X, given Y; [1 mark]

(*iv*) the *mutual information* $i(X; Y)$ between the continuous random variables X and Y; [1 mark]

(*v*) how the *channel capacity* of a continuous channel which takes X as its input and emits Y as its output would be determined. [1 mark]

(*b*) For a time-varying continuous signal $g(t)$ which has Fourier transform $G(k)$, state the *modulation theorem* and explain its rôle in AM radio broadcasting. How does modulation enable many independent signals to be encoded into a common medium for transmission, and then separated out again via tuners upon reception? [4 marks]

(*c*) Briefly define

(*i*) The *Differentiation Theorem* of Fourier analysis: if a function $g(x)$ has Fourier transform $G(k)$, then what is the Fourier transform of the n^{th} derivative of $g(x)$, denoted $g^{(n)}(x)$? [2 marks]

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

(*iii*) If symbols from an alphabet of entropy H are encoded with a *code rate* of R bits per symbol, what is the *efficiency* η of this coding? [2 marks]

(*d*) Briefly explain

(*i*) how 10 V is expressed in dBμV; [1 mark]

(*ii*) the YCrCb coordinate system. [4 marks]

11 Computer Vision

(*a*) Consider the *eigenface* algorithm for face recognition in computer vision.

(*i*) What is the rôle of the database population of example faces upon which this algorithm depends? [3 marks]

(*ii*) What are the features that the algorithm extracts, and how does it compute them? How is any given face represented in terms of the existing population of faces? [4 marks]

(*iii*) What are the strengths and the weaknesses of this type of representation for human faces? What invariances, if any, does this algorithm capture over the factors of perspective angle (or pose), illumination geometry, and facial expression? [4 marks]

(*iv*) Describe the relative computational complexity of this algorithm, its ability to learn over time, and its typical performance in face recognition trials. [3 marks]

(*b*) In a visual inference problem, we have some data set of observed features x, and we have a set of object classes $\{C_k\}$ about which we have some prior knowledge. Bayesian pattern classification asserts that:

$$P(C_k|x) = \frac{P(x|C_k)P(C_k)}{P(x)}$$

Explain the meaning of, and give the name for, each of these three terms:

$P(C_k|x)$
$P(x|C_k)$
$P(C_k)$ [3 marks]

(*c*) Define the concept of *reflectance map* $\phi(i, e, g)$ and define the three variables i, e, and g on which it depends. [3 marks]

12 Numerical Analysis II

The best L_∞ approximation to $f(x) \in C[-1,1]$ by a polynomial $p_{n-1}(x)$ of degree $n-1$ has the property that

$$\max_{x\in[-1,1]} |e(x)|$$

is attained at $n+1$ distinct points $-1 \leq \xi_0 < \xi_1 < \ldots < \xi_n \leq 1$ such that $e(\xi_j) = -e(\xi_{j-1})$ for $j = 1, 2, \ldots n$ where $e(x) = f(x) - p_{n-1}(x)$.

(*a*) Let $f(x) = x^2$. Show, by means of a clearly labelled sketch graph, that the best polynomial approximation of degree 1 is a constant. [3 marks]

(*b*) Now suppose $f(x) = (x+1)/(x+\frac{5}{3})$ is the function to be approximated over $[-1,1]$. By sketching the graph, deduce properties of the best linear approximation $p_1(x)$. By differentiating $e(x)$, find $p_1(x)$. [9 marks]

(*c*) Now consider $f(x) = x/(9x^2+16)$. Explain why the best approximation over $[-1,1]$ of degree 2 or less is of the form $p_2(x) = ax$, and sketch the graph to show the extreme values of $e(x)$. Verify that $x = 4/9$ is one of the extreme values and find a. [8 marks]

13 Specification and Verification I

(*a*) What is partial about partial correctness? [2 marks]

(*b*) What is the difference between a variant and an invariant? [2 marks]

(*c*) Why are annotations needed for mechanising program verification? [2 marks]

(*d*) What additional annotations are needed for total correctness? [2 marks]

(*e*) How do refinement and *post hoc* verification differ? [2 marks]

(*f*) Give an example of a higher-order formula that is not first-order. [2 marks]

(*g*) Why is higher-order logic typed? [2 marks]

(*h*) How are $\{P\}C\{Q\}$ and $\texttt{wlp}(C,Q)$ related? [2 marks]

(*i*) How can $[c]q$ and $<c>q$ be defined in higher-order logic? [2 marks]

(*j*) Explain the difference between soundness and completeness. [2 marks]

14 Natural Language Processing

In (1) and (2) below, the words in the sentences have been assigned tags from the CLAWS 5 tagset by a stochastic part-of-speech (POS) tagger:

(1) Turkey_NP0 will_VM0 keep_VVI for_PRP several_DT0 days_NN2 in_PRP a_AT0 fridge_NN1

(2) We_PNP have_VHB hope_VVB that_CJT the_AT0 next_ORD year_NN1 will_VM0 be_VBI peaceful_AJ0

In sentence (1), *Turkey* is tagged as a proper noun (NP0), but should have been tagged as a singular noun (NN1). In sentence (2), *hope* is tagged as the base form of a verb (VVB: i.e., the present tense form other than for third person singular), but should be NN1. All other tags are correct.

(*a*) Describe how the probabilities of the tags are estimated in a basic stochastic POS tagger. [7 marks]

(*b*) Explain how the probability estimates from the training data could have resulted in the tagging errors seen in (1) and (2). [6 marks]

(*c*) In what ways can better probability estimates be obtained to improve the accuracy of the basic POS tagger you described in part (*a*)? For each improvement you mention, explain whether you might expect it to improve performance on examples (1) and (2). [7 marks]

15 Denotational Semantics

Let D be a domain with bottom element $\bot$. Let $h, k : D \to D$ be continuous functions with h strict (so $h(\bot) = \bot$). Let $\mathbb{B} = \{true, false\}$. Define the conditional function,

$$\text{if} : \mathbb{B}_\bot \times D \times D \to D$$

by $\text{if}(b, d, d') = d$ if $b = true$, d' if $b = false$, and $\bot$ otherwise. Let $p : D \to \mathbb{B}_\bot$ be a continuous function.

The function f is the least continuous function from $D \times D$ to D such that

$$\forall x \in D.\ f(x, y) = \text{if}(p(x),\ y,\ h(f(k(x), y)))\ .$$

(a) State the principle of fixed point induction. What does it mean for a property to be admissible? [4 marks]

(b) Show that

$$\forall b \in \mathbb{B}_\bot, d, d' \in D.\ h(\text{if}(b,\ d,\ d')) = \text{if}(b,\ h(d),\ h(d'))\ .$$

[3 marks]

(c) Prove that the property

$$Q(g) \Leftrightarrow_{def} \forall x, y \in D.\ h(g(x, y)) = g(x, h(y))\ ,$$

where g is a continuous function from $D \times D$ to D, is admissible. [5 marks]

(d) Prove $Q(f)$ by fixed point induction. [8 marks]

16 Computer Systems Modelling

Let X be a random variable taking values in the discrete state space $\{1, 2, \ldots, 6\}$ representing the outcome of a fair die with distribution

$$P(X = i) = \frac{1}{6} \qquad i = 1, 2, \ldots, 6\,.$$

(a) Suppose that you have been given a function `r()` that claims to return pseudo-random numbers with the distribution $U(0, 1)$. Suppose also that you have used this function to generate an independent sample of size 150 of values of X with outcomes given in the following table. Given the relatively high frequency of the outcome $i = 6$ in your sample you may be concerned that the function `r()` is biased. Explain how a goodness of fit test can be used to test for such a bias.

face, i	1	2	3	4	5	6
number of outcomes	22	21	22	27	22	36

[4 marks]

(b) Describe how to apply the χ^2 (Chi-squared) goodness of fit test to your sample.

The following table gives values of t such that if T is a χ^2 random variable with d degrees of freedom then $P(T > t) = 0.05$. Do you conclude from this test that the function `r()` is biased or not at the 5% level?

degrees of freedom d	1	2	3	4	5	6
t	3.84	5.99	7.81	9.49	11.07	12.59

[6 marks]

(c) Suppose that you have implemented a discrete event simulator for a FIFO M/G/1 queueing model for the processing of tasks in a computer system. Explain what probabilistic modelling assumptions are made in specifying such a simulator. Given a log of events and event times generated by your simulator explain what tests you would use to validate your simulation approach.

[10 marks]

END OF PAPER