COMPUTER SCIENCE TRIPOS Part II

Wednesday 2 June 2004 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

1 Comparative Architectures

(*a*) Why is accurate branch prediction so important to modern microprocessor designs? [2 marks]

(*b*) Explain the operation of local history branch predictors, trade-offs made in their design, and limitations in their performance. [6 marks]

(*c*) For what reasons are some branches difficult to predict? What strategies can be employed to mitigate such potential performance problems? [4 marks]

(*d*) Describe a trace cache, and hence explain what advantages it might offer over a traditional instruction cache. [4 marks]

(*e*) It is possible to obtain some of the benefits of a trace cache with a purely software approach, using runtime binary re-writing techniques. How might such a system operate? [4 marks]

2 VLSI Design

(*a*) Sketch the circuit of a dynamic CMOS gate controlled by a clock φ that precharges when $\varphi = 0$ and evaluates the function $\overline{A + B \cdot C}$ when $\varphi = 1$. [4 marks]

(*b*) Explain how your circuit works and describe *two* advantages and *two* disadvantages when compared with static CMOS. [6 marks]

(*c*) Present *two* ways of designing cascaded logic in dynamic CMOS and explain how they work. [8 marks]

(*d*) Present a further modification to the circuit so that its output is retained when the clock stops. [2 marks]

3 Digital Communication II

(*a*) The architecture of IP routers has had to evolve to meet the demands of increasing network link bandwidth. Outline the design of a modern router as might be found in the Internet core. [5 marks]

(*b*) The longest prefix match route lookup function can be accomplished in a number of ways. Describe the operation and relative merits of approaches that involve: [8 marks]

(*i*) a binary trie;

(*ii*) a 2^{24} entry lookup table;

(*iii*) multiple hash tables.

(*c*) Explain the operation of a virtual output buffered switch fabric, and hence explain its advantages over input buffered and output buffered designs. [7 marks]

4 Distributed Systems

(*a*) Define strong and weak consistency. [2 marks]

(*b*) A process group manages a set of widely distributed replicas. The group is open and unstructured; that is, external processes may invoke any group member for reading or writing.

(*i*) Discuss how the replicas can be kept strongly consistent in the presence of concurrent invocations and failures. [12 marks]

(*ii*) Would it be more appropriate to use a structured group (with a single co-ordinator) to manage the replicas? Justify your answer. [6 marks]

Your solution should discuss the *selection and use* of algorithms and protocols. It is not necessary to specify them in great detail.

5 Advanced Systems Topics

A *distributed shared virtual memory* (DSVM) programming model is often used on cluster computers because it can allow multi-threaded applications to be distributed across a set of machines without needing to be re-written.

(*a*) Describe the implementation of DSVM using a centralized page manager. Your answer should identify:

(*i*) What data structures are maintained by the page manager.

(*ii*) What happens when a machine performs a read operation to a page.

(*iii*) What happens when a machine performs a write operation to a page. [8 marks]

(*b*) Someone observes that the centralized page manager may form a *bottleneck* and a *single point of failure*. Do you agree with these observations? [2 marks]

(*c*) Sketch the implementation of a scalable spin-lock for use on shared-memory multiprocessor machines. You may assume the existence of an atomic *compare-and-swap* operation. [5 marks]

(*d*) Do you think that your spin-lock design would be appropriate for use on a DSVM system? Either explain why it will perform well, or suggest an alternative implementation which would be appropriate. [5 marks]

6 Security

A car locking system consists of an engine management system E which shares a key K with a microcontroller M embedded in the key fob. When an attempt is made to open the door, a challenge N_C is sent by E to M; M computes a response N_R by encrypting N_C with K using a block cipher.

$$E \to M \ : \ N_C$$
$$M \to E \ : \ N_R = \{N_C\}_K$$

M must respond to a challenge in 100 ms, which means that the total length of N_C and N_R can be no more than 64 bits. In addition, if a wrong response is received, E will wait 900 ms before sending another challenge, so that only one trial response can be attempted per second.

A hotel parking valet has access to guests' keys for a few hours or days at a time. He builds test equipment to try out many random challenges and thus constructs a table of (N_C, N_R) pairs. His goal is to follow a guest home, try to unlock the door until E sends a challenge N_C already in the table, whereupon he will return the corresponding N_R and steal the car.

(a) Which is the most secure design against this type of attack – $N_C = 24$ bits and $N_R = 40$ bits, $N_C = N_R = 32$ bits, or $N_C = 40$ bits and $N_R = 24$ bits? Justify your answer. [5 marks]

(b) Is it important whether the underlying block cipher is AES or DES? Justify your answer. [5 marks]

(c) Such a design has been fielded and a long-term contract awarded for the manufacture of key fobs. As a consequence, only the engine controller software can be modified. Is there a modification that makes the valet attack significantly harder? [5 marks]

(d) The company that owns the patent on this protocol now wishes to sell handheld password generators, containing the key-fob chips, to banks, with a view to authenticating their customers and thus stopping "phishing" attacks. If you were a bank security manager, would you be enthusiastic about such a proposed solution? [5 marks]

7 Optimising Compilers

(*a*) Explain the concept of "strength reduction" when applied to loops, illustrating your explanation with the C loop

```
extern int a[M][N];
for (i=0; i<M; i++) for (j=0; j<N; j++) t += a[i][j];
```

Consider the issue of whether strength reduction is sufficient to reduce this to a single loop. [8 marks]

(*b*) The "loop invariant lifting" optimisation says that if an expression is *available* at a point within a loop, and none of its free variables may be updated within the loop, then the expression may be instead computed just before entry to the first iteration of the loop.

Explain the concepts "available" and "computed just before entry" in more detail, focusing your argument by explaining how the following C program could be optimised, expressing the resulting optimised code in a similar syntax to the original.

```
extern int u[100],v[100],w[100];
void f(int n)
{   int i, y = ..., z = ...;
    for (i=5; i<n; i++)
    {    u[i] += 1000/y;
         v[i] += 1000/z;
         p(&y);
         w[i] += 1000/z;
    }
}
```

You need not consider "strength reduction" optimisations here and, because you are expressing the resultant code in C, there is no need to consider register allocation issues. [12 marks]

8 Artificial Intelligence

(*a*) Describe the way in which a problem should be represented in order to allow its solution using a *heuristic search* technique. [5 marks]

(*b*) Define what it means for a search algorithm to be *complete*, and to be *optimal*. [2 marks]

(*c*) Define what it means for a heuristic function to be *admissible*, and to be *monotonic*. [2 marks]

(*d*) Describe the operation of the $A^\star$ heuristic search algorithm. [5 marks]

(*e*) Prove that the $A^\star$ heuristic search algorithm is optimal when applied in conjunction with a monotonic heuristic. State the conditions under which the algorithm is also complete, and explain why this is the case. [6 marks]

9 Database Theory

(*a*) Define formally the semi-structured data (SSD) model. [5 marks]

(*b*) Show how SSD expressions can be expressed in XML. [2 marks]

(*c*) What are the main differences between the SSD and XML models? [2 marks]

When viewed graphically, simple SSD expressions denote trees. Consider a variant, d-SSD, where the edges emanating from any node in the tree must have a *unique* label, but where the labels may themselves be d-SSD expressions. (You may disregard object identities (oids); hence d-SSD expressions always denote trees.)

(*d*) Define the syntax of d-SSD expressions. [3 marks]

(*e*) Give a d-SSD expression to represent the following:

(*i*) the array [“Do”,“Re”,“Mi”];

(*ii*) the set $\{11, 52, 44\}$;

(*iii*) the bag $\{\{10, 10, 13, 42, 13, 10\}\}$. [1 mark each]

(*f*) Define the syntax for path expressions in the d-SSD model. [2 marks]

(*g*) Hence describe precisely how the d-SSD model can be extended to represent graphs. (Answers that use oids will receive little credit.) [3 marks]

10 Information Theory and Coding

(*a*) Consider a binary symmetric communication channel, whose input source is the alphabet $X = \{0, 1\}$ with probabilities $\{0.5, 0.5\}$; whose output alphabet is $Y = \{0, 1\}$; and whose channel matrix is

$$\begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}$$

where ϵ is the probability of transmission error.

(*i*) What is the entropy of the source, $H(X)$? [2 marks]

(*ii*) What is the probability distribution of the outputs, $p(Y)$, and the entropy of this output distribution, $H(Y)$? [2 marks]

(*iii*) What is the joint probability distribution for the source and the output, $p(X, Y)$, and what is the joint entropy, $H(X, Y)$? [2 marks]

(*iv*) What is the mutual information of this channel, $I(X; Y)$? [2 marks]

(*v*) How many values are there for ϵ for which the mutual information of this channel is maximal? What are those values, and what then is the capacity of such a channel in bits? [2 marks]

(*vi*) For what value of ϵ is the capacity of this channel minimal? What is the channel capacity in that case? [2 marks]

(*b*) A variable length, uniquely decodable code which has the prefix property, and whose N binary code word lengths are $n_1 \leq n_2 \leq n_3 \leq \cdots \leq n_N$ must satisfy what condition on code word lengths? (State the condition, and name it.) [3 marks]

(*c*) You are asked to compress a collection of files, each of which contains several thousand photographic images. All images in a single file show the same scene. Everything in this scene is static (no motion, same camera position, etc.) except for the intensity of the five light sources that illuminate everything. The intensity of each of the five light sources changes in completely unpredictable and uncorrelated ways from image to image. The intensity of each pixel across all photos in a file can be described as a linear combination of the intensity of these five light sources.

(*i*) Which one of the five techniques *discrete cosine transform*, *μ-law coding*, *2-D Gabor transform*, *Karhunen-Loève transform* and *Golomb coding* would be best suited to remove redundancy from these files, assuming your computer is powerful enough for each? [1 mark]

(*ii*) Explain briefly this transform and why it is of use here. [4 marks]

11 Computer Vision

(a) When defining and selecting which features to extract in a pattern classification problem, what is the goal for the statistical clustering behaviour of the data in terms of the variances within and amongst the different classes? [2 marks]

(b) Consider the following pair of filter kernels:

-1	-1	-1	-1	-1	-1
-1	-3	-4	-4	-3	-1
2	4	5	5	4	2
2	4	5	5	4	2
-1	-3	-4	-4	-3	-1
-1	-1	-1	-1	-1	-1

1	1	1	1	1	1
-1	-2	-3	-3	-2	-1
-1	-3	-4	-4	-3	-1
1	3	4	4	3	1
1	2	3	3	2	1
-1	-1	-1	-1	-1	-1

(i) Why do these kernels form a quadrature pair? [2 marks]

(ii) What is the "DC" response of each of the kernels, and what is the significance of this? [1 mark]

(iii) To which orientations and to what kinds of image structure are these filters most sensitive? [1 mark]

(iv) Mechanically how would these kernels be applied directly to an image for filtering or feature extraction? [1 mark]

(v) How could their respective Fourier Transforms alternatively be applied to an image, to achieve the same effect as in (iv)? [1 mark]

(vi) How could these kernels be combined to locate facial features? [2 marks]

(c) Explain why inferring object surface properties from image properties is, in general, an ill-posed problem. In the case of inferring the colours of objects from images of the objects, how does knowledge of the properties of the illuminant affect the status of the problem and its solubility? [5 marks]

(d) Explain and illustrate the "Paradox of Cognitive Penetrance" as it relates to computer vision algorithms that we know how to construct, compared with the algorithms underlying human visual competence. Discuss how human visual illusions may relate to this paradox. Comment on the significance of this paradox for computer vision research. [5 marks]

12 Numerical Analysis II

(*a*) State a recurrence formula for the sequence of Chebyshev polynomials, $\{T_k(x)\}$, and list these as far as $T_5(x)$. [4 marks]

(*b*) What is the best L_∞ polynomial approximation over $[-1, 1]$ to x^k using polynomials of lower degree, and what is its degree? Use this property to explain the method of economisation of a Taylor series. How can the error in one economisation step be estimated? [7 marks]

(*c*) It is required to approximate the function $f(x) = \lim_{k\to\infty} P_k(x)$ over $[-1, 1]$ with an absolute accuracy of 2 decimal places, where

$$P_k(x) = \sum_{n=1}^{k} \frac{x^n}{n\ n!} \ .$$

As this series converges faster than e^x, a good estimate of the error $||f(x) - P_k(x)||_\infty$ in the truncated Taylor series is given by evaluating the next term

$$\frac{x^{k+1}}{(k+1)(k+1)!}$$

at $x = 1$. Use the method of economisation to find a polynomial approximation of the required accuracy. [9 marks]

13 Specification and Verification I

(*a*) Describe how to use Floyd–Hoare logic to specify that a program sorts an array A so that $A(0), A(1), \ldots, A(N)$ are in ascending order. [8 marks]

(*b*) What is VDM notation? Use the sorting example to show how it can shorten specifications. Does the VDM specification of sorting have the same meaning as the Floyd–Hoare specification you gave in answer to part (*a*)? [6 marks]

(*c*) Describe the concept of weakest preconditions and weakest liberal preconditions. How do they relate to Floyd–Hoare specifications? [6 marks]

14 Natural Language Processing

A context free grammar for a fragment of English is shown below:

```
S -> NP VP
NP -> Det N
N -> N N
VP -> rumbles, rusts
Det -> the, a, every
N -> bus, car, train, park, airport, station
```

(*a*) Show the parse trees for the two parses that the grammar assigns for sentence S1.

S1: the train station bus rumbles

[3 marks]

(*b*) Give an algorithm for a bottom-up passive chart parser without packing. Illustrate your answer by showing the edges constructed when parsing sentence S1. [11 marks]

(*c*) Describe how this algorithm could be modified so that edges may be *packed*, illustrating your answer by considering sentences S1 and S2. What effect does packing have on parsing efficiency?

S2: the airport car park bus rumbles

[6 marks]

15 Denotational Semantics

(*a*) The function *fix* is the least fixed point operator from $(D \to D)$ to D, for a domain D.

(*i*) Show that $\lambda f.\ f^n(\bot)$ is a continuous function from $(D \to D)$ to D for any natural number n.
[Hint: Use induction on n. You may assume the evaluation function $(f, d) \mapsto f(d)$ and the function $f \mapsto (f, f)$, where $f \in (D \to D)$ and $d \in D$, are continuous.] [7 marks]

(*ii*) Now argue briefly why

$$\mathit{fix} = \bigsqcup_{n \geq 0} \lambda f.\ f^n(\bot)\ ,$$

to deduce that *fix* is itself a continuous function. [3 marks]

(*b*) In this part you are asked to consider a variant $\mathbf{PCF_{rec}}$ of the programming language **PCF** in which there are terms $\mathbf{rec}\, x : \tau.\, t$, recursively defining x to be t, instead of terms $\mathbf{fix}_\tau$.

(*i*) Write down a typing rule for $\mathbf{rec}\, x : \tau.\, t$. [2 marks]

(*ii*) Write down a rule for the evaluation of $\mathbf{rec}\, x : \tau.\, t$. [2 marks]

(*iii*) Write down the clause in the denotational semantics which describes the denotation of $\mathbf{rec}\, x : \tau.\, t$. (This will involve the denotation of t which you may assume.) [3 marks]

(*iv*) Write down a term in $\mathbf{PCF_{rec}}$ whose denotation is the least fixed point operator of type $(\tau \to \tau) \to \tau$. [3 marks]

16 Computer Systems Modelling

Suppose that bus inter-arrival times, X, at a given bus stop have a probability density function $f_X(x)$ with mean $\mu = E(X)$ and variance $\sigma^2 = \mathrm{Var}(X) = E(X^2) - \mu^2$. Suppose that a randomly arriving customer arrives during a bus inter-arrival interval of length Y and suppose that the probability density of Y is $f_Y(y)$. It may be assumed that

$$f_Y(y) = Cyf_X(y)$$

for some constant C.

(a) Derive an expression for the constant C in terms of μ and σ^2. [7 marks]

(b) Derive an expression for the average waiting time as seen by a randomly arriving customer. [7 marks]

(c) For each of the following cases, calculate the average waiting time as seen by a randomly arriving customer.

(i) X is deterministic taking a value of 10. [2 marks]

(ii) X is exponentially distributed with mean $\mu = 10$. [2 marks]

(iii) X has a general distribution with mean $\mu = 10$ and variance $\sigma^2 = 500$. [2 marks]

END OF PAPER