COMPUTER SCIENCE TRIPOS Part II (General)
DIPLOMA IN COMPUTER SCIENCE

Thursday 3 June 2004 1.30 to 4.30

Paper 13 (Paper 4 of Diploma in Computer Science)

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 Data Structures and Algorithms

(*a*) Describe an efficient algorithm to determine whether two finite line segments in a plane intersect. You may assume that the end points of each line are given as x–y coordinates. [8 marks]

(*b*) Describe in detail an efficient algorithm to find the convex hull of a set of points lying on a plane. Show that the complexity of the Graham scan used in the algorithm is $O(n)$ and that the algorithm as a whole has complexity $O(n \log n)$. [8 marks]

(*c*) Discuss how it is possible to eliminate many of the points before the convex hull algorithm is entered. [4 marks]

2 Computer Design

The ARM processor allows the second operand to be shifted by an arbitrary amount. In order to improve the performance, a six-stage pipeline is proposed with the following stages:

instruction fetch	decode and register fetch	shift operand 2	execute	memory access	register write back

(*a*) What are control hazards and how could they be resolved in the above pipeline? [4 marks]

(*b*) What are data hazards and how could they be resolved in the above pipeline? [4 marks]

(*c*) What are feed-forward paths and where could they be added to the above pipeline to improve performance? [6 marks]

(*d*) Why might a branch instruction result in pipeline bubbles and how many bubbles will appear in the above pipeline as a result of taking a branch instruction? [6 marks]

3 Digital Communication I

(*a*) Define the terms *latency* and *capacity* as applied to communication channels. [2 marks]

(*b*) Is there a strict relation between the two? [1 mark]

(*c*) Show how the latency of a channel can have a direct effect on the capacity of a higher-layer channel which uses it. [10 marks]

(*d*) How can the capacity of the higher-layer channel be improved (keeping the characteristics of the underlying channel unchanged)? [4 marks]

(*e*) In what circumstances might these improvements have only limited benefit? [3 marks]

4 Distributed Systems

A network-based service manages persistent objects. The service must enforce an access control policy to protect the objects.

(*a*) Discuss how this access control might best be implemented for the following example of objects and policy components:

> Objects: Files in a University Department's file service, operating behind a firewall.
>
> Policy: The owner may specify read, write and execute rights in terms of principals and groups. [4 marks]

(*b*) Discuss how this access control might best be implemented for *two* of the following examples:

(*i*) Objects: Files in a commercial, distributed, Internet-based file service.
Policy: The owner may authorise other principals to download the file.

(*ii*) Objects: Sales data relating to a company.
Policy: Those employed in the Sales Departments of all branches of the company worldwide may read the data.

(*iii*) Objects: Electronic health records (EHRs) in a nationwide service.
Policy: The owner (patient) may read from its own EHR. A qualified and employed doctor may read and write the EHR of a patient registered with him/her.

(*iv*) Object: The solution to online coursework.
Policy: The coursework setter has read and write access. A candidate has no access until after the marks have been published.

[8 marks each]

5 Computer Graphics and Image Processing

(*a*) Explain why display devices appear to be able to reproduce (almost) all the colours of the spectrum using only red, green and blue light. [4 marks]

(*b*) Describe an algorithm (other than thresholding) which will convert a greyscale image (8 bits per pixel) to a bi-level black and white image (1 bit per pixel), with the same number of pixels, while retaining as much detail as possible. [8 marks]

(*c*) Explain what specular and diffuse reflection are in the real world. State and explain equations for calculating approximations to both in a computer. [8 marks]

6 Compiler Construction

(a) Explain the differences (illustrating each with a small program) between

(i) static and dynamic binding (scoping); [4 marks]

(ii) static and dynamic typing. [2 marks]

(b) Java is sometimes said to be "dynamically typed" in that a variable whose type is (class) C can be assigned a value of (class) D provided that D extends C; conversely a variable of type D can be assigned a value of type C using a cast. By considering storage layouts, explain why the former assignment is always valid and the latter *sometimes* invalid. [4 marks]

(c) A new programming language has the notion of "statically scoped exceptions" in which the program

```
exception foo;
void f()
{   try
    {   void g() { raise foo; }
        try {
          g();
        }
        except (foo) { C2 }
    }
    except (foo) { C1 }
}
```

would execute `C1` rather than `C2` as the former was in scope at the `raise` point. By analogy with statically scoped variables, or otherwise, explain how such exceptions might be implemented on a stack. [10 marks]

7 Artificial Intelligence

In the following, N is a feedforward neural network architecture taking a vector

$$\mathbf{x}^T = (\ x_1 \quad x_2 \quad \cdots \quad x_n\ )$$

of n inputs. The complete collection of weights for the network is denoted $\mathbf{w}$ and the output produced by the network when applied to input $\mathbf{x}$ using weights $\mathbf{w}$ is denoted $N(\mathbf{w}, \mathbf{x})$. The number of outputs is arbitrary. We have a sequence $\mathbf{s}$ of m labelled training examples

$$\mathbf{s} = ((\mathbf{x}_1, \mathbf{l}_1), (\mathbf{x}_2, \mathbf{l}_2), \ldots, (\mathbf{x}_m, \mathbf{l}_m))$$

where the $\mathbf{l}_i$ denote vectors of desired outputs. Let $E(\mathbf{w}; (\mathbf{x}_i, \mathbf{l}_i))$ denote some measure of the error that N makes when applied to the ith labelled training example. Assuming that each node in the network computes a weighted summation of its inputs, followed by an activation function, such that the node j in the network computes a function

$$g\left(w_0^{(j)} + \sum_{i=1}^{k} w_i^{(j)} \mathrm{input}(i)\right)$$

of its k inputs, where g is some activation function, derive in full the backpropagation algorithm for calculating the gradient

$$\frac{\partial E}{\partial \mathbf{w}} = \left(\ \frac{\partial E}{\partial w_1} \quad \frac{\partial E}{\partial w_2} \quad \cdots \quad \frac{\partial E}{\partial w_W}\ \right)^T$$

for the ith labelled example, where $w_1, \ldots, w_W$ denotes the complete collection of W weights in the network.

[20 marks]

8 Databases

(a) Define the operators of the core relational algebra. [5 marks]

(b) Let R be a relation with schema $(A_1, \ldots, A_n, B_1, \ldots, B_m)$ and S be a relation with schema $(B_1, \ldots, B_m)$. The quotient of R and S, written $R \div S$, is the set of tuples t over attributes $(A_1, \ldots, A_n)$ such that for every tuple s in S, the tuple ts (i.e. the concatenation of tuples t and s) is a member of R. Define the quotient operator using the operators of the core relational algebra. [8 marks]

(c) The core relational algebra can be extended with a duplicate elimination operator, and a grouping operator.

(i) Define carefully these two operators. [3 marks]

(ii) Assuming the grouping operator, show how the duplicate elimination operator is, in fact, unnecessary. [2 marks]

(iii) Can the grouping operator be used to define the projection operator? Justify your answer. [2 marks]

9 Numerical Analysis II

(*a*) A Riemann integral over $[a, b]$ is defined by

$$\int_a^b f(x)\,dx = \lim_{\substack{n\to\infty \\ \Delta\xi\to 0}} \sum_{i=1}^{n} (\xi_i - \xi_{i-1}) f(x_i)\ .$$

Explain the terms *Riemann sum* and *mesh norm*. [4 marks]

(*b*) Consider the quadrature rule

$$Qf = \frac{3h}{8}[f(a) + 3f(a+h) + 3f(a+2h) + f(a+3h)] - \frac{3f^{(4)}(\lambda)h^5}{80}\ .$$

If $[a, b] = [-1, 1]$ find $\xi_0, \xi_1, \ldots, \xi_4$ and hence show that this is a Riemann sum. [3 marks]

(*c*) Suppose R is a rule that integrates constants exactly over $[-1, 1]$, and that $f(x)$ is bounded and Riemann-integrable over $[a, b]$. Write down a formula for the composite rule $(n \times R)f$ and prove that

$$\lim_{n\to\infty} (n \times R)f = \int_a^b f(x)\,dx$$ [6 marks]

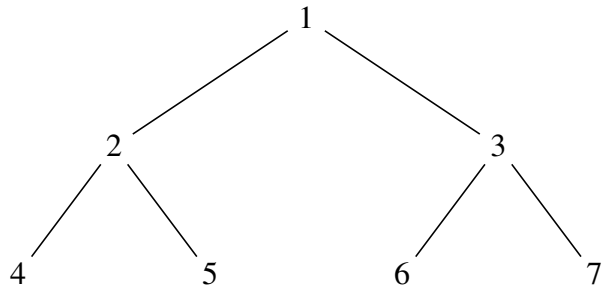
(*d*) What is the formula for $(n \times Q)f$ over $[a, b]$? [4 marks]

(*e*) Which polynomials are integrated exactly by Qf? Which monomials are integrated exactly by the product rule $(Q \times Q)F$ when applied to a function of x and y? [3 marks]

10 Introduction to Functional Programming

(a) Define a polymorphic datatype `'a tree` to represent binary trees. [1 mark]

(b) A *breadth-first* traversal of a tree walks over all the nodes at each level before proceeding to the next level. For example the breadth-first traversal of the tree:



visits the nodes in the order $1, 2, 3, 4, 5, 6, 7$.

Define a function `breadth: 'a tree -> 'a list` such that `breadth(t)` returns the nodes of tree `t` in breadth-first order. [10 marks]

(c) Define a polymorphic datatype `'a seq` to represent lazy lists. [1 mark]

(d) Define a polymorphic datatype `'a ltree` to represent lazy binary trees. [3 marks]

(e) Define a function `inorder` of type `'a ltree -> 'a seq` that traverses a lazy tree in-order, returning the nodes in a lazy list. (You should define any auxiliary functions you may use.) [5 marks]

11 Natural Language Processing

(*a*) Give brief definitions of the following terms:

(*i*) referring expression;

(*ii*) cataphora;

(*iii*) pleonastic pronoun.

[6 marks]

(*b*) Describe the Lappin and Leass algorithm for pronoun resolution, illustrating its operation on the text below. Exact weights for salience factors are *not* required.

> Owners love the new hybrid cars. They all say that they have much better fuel economy than conventional vehicles. And it seems that the performance of hybrid cars matches all expectations.

[14 marks]

12 Complexity Theory

(*a*) Define a one-way function. [4 marks]

(*b*) Explain why the existence of one-way functions would imply that P≠NP. [7 marks]

(*c*) Recall that `Reach` is the problem of deciding, given a graph G a source vertex s and a target vertex t, whether G contains a path from s to t; and `Sat` is the problem of deciding whether a given Boolean formula is satisfiable.

For each of the following statements, state whether it is true or false and justify your answer.

(*i*) If `Reach` is NP-complete then P=NP. [3 marks]

(*ii*) If `Reach` is NP-complete then NP≠PSPACE. [3 marks]

(*iii*) If `Sat` is PSPACE-complete then NP=PSPACE. [3 marks]

END OF PAPER