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

Thursday 4 June 1998 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.*
Write on **one** *side of the paper only.*

1 Distributed Systems

Discuss capability-based access control under the headings

- protection of capabilities
- control of transfer of capabilities between principals
- delegation of rights
- revocation of capabilities

Your answer should mention the differences between the management of capabilities in distributed as opposed to centralised systems. You should consider alternative designs of capabilities. [20 marks]

2 Computer Design

Computer memory is usually organised as a *memory hierarchy*. Why is this the case? [4 marks]

What are the relative latency and bandwidth characteristics of each level of a typical memory hierarchy? [4 marks]

How does a direct-mapped cache work and what might the data replacement policy be? [6 marks]

How does a set-associative cache work and what might the data replacement policy be? [6 marks]

3 Digital Communication I

You are required to design a topology discovery protocol for a network of switching nodes interconnected by links. There are n nodes, l links, the maximum degree of any node is k and there is a path between any two nodes of not more than d hops. All links are bi-directional.

Each node has a unique identifier of four bytes which it knows.

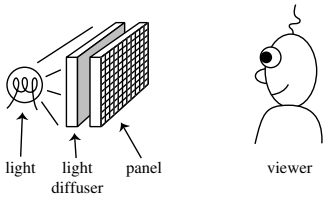
(a) Design a protocol (including message formats) for a node to learn about its immediate neighbours. [5 marks]

(b) Design a protocol (including message formats) for distributing this information across the network. [10 marks]

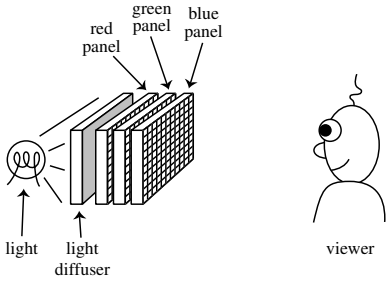
(c) Give a bound on the total amount of information which is transmitted to ensure that every node acquires complete topology information. [5 marks]

4 Computer Graphics and Image Processing

An inventor has recently developed a new display device: it is a transparent panel with a rectangular array of square pixels. The panel is tinted with a special ink which allows each pixel to range from totally transparent to transmitting only the colour of the ink. Each pixel has an 8-bit value. For example, if the ink is blue then a pixel value of 0 would be totally transparent, 255 totally blue (only blue light transmitted) and 100 a light blue.



The inventor has recently found that he can make the special ink in *any* colour he likes, but that each panel can be tinted with only one of these colours. He proposes to use three inks in three panels to make a 24-bit colour display: a red-tinted panel, a green-tinted panel and a blue-tinted panel will be stacked up to make a full-colour display (see picture). A value of $(0, 0, 0)$ will thus be white (transparent), $(255, 0, 0)$ red and $(255, 255, 255)$ black.



Explain why this will not work. [4 marks]

Modify the three-panel design so that it will work. [3 marks]

In common with other 24-bit "full-colour" displays (for example CRT, LCD), your display *cannot* display *every* colour which a human can perceive. Why not? [3 marks]

In image compression we utilise three different mechanisms to compress pixel data:

(a) mapping the pixel values to some other set of values

(b) quantising those values

(c) symbol encoding the resulting values

Explain each mechanism, why it helps us to compress the image, and whether (giving reasons) the resulting image noticeably differs. [10 marks]

5 Business Studies

You are assigned the role of UK sales and marketing manager for a new kind of low-cost computer, primarily aimed at the educational market. Whilst not directly PC compatible, the computer includes web access, PC-compatible word processing, other PC-compatible productivity tools, and a suite of educational programs.

How would you approach this task? Draw up an outline business plan as follows.

(a) Show what communication and distribution channels you propose. [5 marks]

(b) Propose a selling price, and estimate the number of units you might sell at this price. [5 marks]

(c) Estimate a 3-year budget for the sales and marketing activity. [5 marks]

(d) State how you would refine your estimates, and what monitoring you would put in place. [5 marks]

Background information:

- There are about 32,000 schools in the UK.
- The UK government has recently published a consultative document *The National Grid for Learning* with a proposal to spend £100M on provision of internet access in schools over the next 5 years. This sum includes infrastructure provision, content development and teacher training, as well as a contribution to provision of computers in schools.
- The average school IT spend is projected to be £18,000 each year. Additional funding may be available from government and parents for specific projects.
- The unit manufacturing cost is £200, delivered.

6 Compiler Construction

You have been given a new programming language with a C-like syntax, with integer variables and functions and with static binding of free variables. Your manager can parameterise certain aspects of the language, including the following three options:

- For "`int x = e;`" whether the variable `x` has the same l-value of `e` or whether a new l-value is created and initialised to the r-value of `e`. If `e` is only an r-value then a new l-value is created in both circumstances.

- For "`int f(int x) { ... }`" whether the variable `x` is passed by l-value ("by reference") or by r-value ("by value"). If the switch is set to "l-value" and the value passed is only an r-value then a new l-value is created, initialised and passed.

- For "`int f(int x) { ... y ... }`" (where the variable `y` is free to `f`) whether the value of `y` is calculated at the times of its uses (association by l-value) or at the time of the definition of `f` (association by r-value).

As a test of your programming skills your manager asks you to write a program which tells how the language has been parameterised. Do so by printing a 3-digit decimal number where the "hundreds" digit is one or two according to whether the first option is by l-value or r-value respectively, similarly with the "tens" digit for the second and with the "units" digit for the third option.

[10 marks]

Explain the structure of an object module which an assembler or compiler might produce to be processed by a linker. Your answer should include discussion of the various object module features needed to represent the compiled form of the C program:

```
int a[10] = { 2,3,5,7,11,13,17,19,23,29 };
extern int b[10];
extern int g(int);
int f(int y)
{   return g(y) + b[5] + a[6];
}
```

[10 marks]

7 Comparative Programming Languages

Many languages either forbid explicit pointer arithmetic or restrict its use. What kinds of problem are they seeking to avoid? Why is it allowed (with some restrictions) in C and C++, and commonly used by programmers? [8 marks]

What are the dangers inherent in allowing memory deallocation to be under the direct control of the programmer? Given these dangers, why has memory deallocation not been automated in standardised C or C++? [12 marks]

8 Prolog for Artificial Intelligence

According to the rules of the Billy Badger Fan Club, an applicant is acceptable for membership provided that:

- The applicant must have two proposers who are members of the club.
- The applicant must be aged between 18 and 30 years of age (inclusive).
- Each proposer must have been a member for at least two years.
- Each proposer must not be a parent of the applicant.

Write a Prolog program that includes a rule for deciding whether an applicant is acceptable for membership, illustrating with a sample database. [20 marks]

9 Databases

Explain what is meant by a *referential integrity constraint* in a relational data model. [4 marks]

The University of Cambridge is determined to maintain its standards under increasing financial pressure. The government maintains league tables of various kinds: teaching quality, research rating, unit cost of each student place. The university still enjoys a high reputation worldwide, but it is in competition with institutions such as MIT and Stanford whose unit costs are much greater. The only way to provide facilities such as new research laboratories, graduate accommodation, a much-needed swimming pool, is by public appeal. It is vital that there is close liaison between the university development offices and colleges so that prospective donors do not become alienated by simultaneous demands.

You have been invited by the university to assist in recording details of its fund-raising. The main purpose is to coordinate the activities of *agencies* such as the university and college development offices in their dealings with *prospects*. The latter may be alumni of some college, charitable foundations or major companies with an educational commitment; it is important to record their *interests* (bioscience, student welfare, sport) so that they can be approached in a favourable context; also their potential resources, in order to maximise the possible benefit to the university. At any time a number of *projects* need funding: each will be developed by a single agency, requiring that a target sum be raised by a given date; projects will have one or more *purposes* which may be linked with the interests of prospects. The other main use of the database is to keep a *diary* of interaction with prospects; in order to retain their goodwill it is essential to know who has been invited where and when, and in what context.

Design the schema for a relational database that will record this information. State clearly any assumptions that you need to make in order to complete the design, and indicate any difficulties that you foresee in maintaining the database.

[16 marks]

10 Introduction to Functional Programming

The following datatypes are meant to be used to represent programs written in a simple imperative language:

```
datatype expression = Expr of (string -> int) -> int;
datatype command = Assign of (string * expression)
                 | Sequence of (command * command)
                 | While_do of (expression * command);
```

The *state*, namely the values of the variables at a given point, is represented by a function that takes the variable name and gives the corresponding value. Variables and expressions only involve integers and, when treated as booleans, zero is regarded as *false* and non-zero values as *true*. For example, the compound command:

```
x := 1;
while n <> 0 do
 { x := x * n;
   n := n - 1 }
```

can be written

```
Sequence(Assign("x",Expr(fn s => 1)),
    While_do(Expr(fn s => s"n"),
        Sequence(Assign("x",Expr(fn s => s"x" * s"n")),
                 Assign("n",Expr(fn s => s"n" - 1)))));
```

A command can be interpreted as a mapping from the initial state to the final state; this will be achieved by the function `interpret` below.

First write an ML function `update` whose type is

```
(string -> int) * string * int -> (string -> int)
```

such that `update(s,x,i)` gives a new state representing state `s` but with the variable `x` being assigned the value `i`. [5 marks]

Now, using `update` or otherwise, write an ML function `interpret` whose type is

```
command -> (string -> int) -> (string -> int)
```

which takes a command c and an initial state s_1, and returns the corresponding final state s_2 that results from executing c in state s_1. For example, if the above compound command is bound to `fp`, then

```
interpret fp (fn "n" => 6) "x";
```

should yield 720. [15 marks]

11 Computer Vision

It could be said that the central problem of pattern recognition is the relation between the within-class variability and the between-class variability for the patterns that one would like to recognise. Explain this problem in the case of face recognition, treating separately the problems of

(*a*) *face detection* (distinguishing faces from non-faces)

(*b*) *face identification*

(*c*) *face interpretation* (classifying the expression and pose angle of the face)

How do the forms of variability for faces influence each of the three tasks? Is within-class variability ever helpful, and between-class variability ever harmful, to the performance of the task? What role can statistical decision-theory play in formalising and solving these problems?

[20 marks]

12 Complexity Theory

Here are some informally expressed opinions about computational complexity. They may be correct, incorrect, misleading or meaningless. In some cases the truth or otherwise of the statement might not be known, either in the sense of it not having been covered in the course or by the answer not being known by anybody anywhere. For each statement comment on its validity and in cases where that is both necessary and straightforward produce an adjusted version of the observation that is properly valid. You are not expected to include proofs to support your claims.

(a) Problems that are not NP-complete are easy to solve. [3 marks]

(b) Problems that are NP-complete will never be solved in reasonable amounts of time even though computers continue to get faster and faster. [3 marks]

(c) To test a number N to see whether it is prime you just have to do a test-division by each of the numbers from 2 to $N-1$, and since there are only $N-2$ of these and division can be done in time $O(n^2)$ this is polynomial time. Thus primality testing is in the class P. [4 marks]

(d) There is a polynomial-time reduction from the k-clique problem to 3-SAT. [3 marks]

(e) There is a polynomial-time reduction from 3-SAT to the k-clique problem. [3 marks]

(f) There have been proposals that biological computers based on DNA might use the massive parallelism of their biochemical activity to solve NP problems rapidly. If such systems could be made to work reliably this would solve the theoretical challenge posed by the concept of NP-completeness. [4 marks]

13 Numerical Analysis II

Explain what is meant by *local error* and *global error* in methods for the solution of *ordinary differential equations* (*ODEs*). If a typical method has local error $O(h^3)$, what would you expect the global error to be? [3 marks]

Euler's method for solution of $y' = f(x, y)$ can be expressed as $y_{n+1} = y_n + k_1$. From the Taylor series, find an expression for k_1. [2 marks]

The Runge–Kutta method RK2 is

$$y_{n+1} = y_n + \tfrac{1}{2}(k_1 + k_2)$$

where k_1 is the increment used by Euler's method, and

$$k_2 = h\, f(x_n + h, y_n + k_1).$$

In terms of Euler's method, what does the quantity k_2 represent? [2 marks]

Assume that RK2 is carried out with step sizes h and $h/2$, and that

$$y_{(h)}(x_{n+1}) = y(x_{n+1}) + C_n h^2 + O(h^3).$$

Derive an estimate of the error $E_n = |C_n|(h/2)^2$ in $y_{(h/2)}(x_{n+1})$. [3 marks]

Let ε be the *target error per unit step*. Why, in *step-size control* for RK2, is $\varepsilon' = \varepsilon/8$ taken as the target error corresponding to half the step size? [2 marks]

A certain ODE is to be solved using RK2 with step-size control. Using computed values for y from the table below, taking $\varepsilon = 0.005$, and starting with $h = 0.1$, state at which values of x you would make the *first* and *second* changes of step size, and what new values of h you would use in each case.

		h			
		0.025	0.05	0.1	0.2
x	0.05	0.10038	0.10050		
	0.1	0.20279	0.20304	0.20400	
	0.15	0.30946	0.30981		
	0.2	0.42295	0.42341	0.42516	0.43200
	0.25	0.54649	0.54702		
	0.3	0.68434	0.68490	0.68697	
	0.35	0.84247	0.84295		
	0.4	1.02971	1.02989	1.03047	1.03373
	0.45	1.25995	1.25930		
	0.5	1.55646	1.55379	1.54484	

[8 marks]