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

Tuesday 6 June 2000 1.30 to 4.30

Paper 11 (Paper 2 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 Digital Electronics

What is the maximum number of terms there can be in a minimal sum of products form of a function of n boolean variables? [2 marks]

Consider a two-bit multiplier with inputs x_1, x_0, y_1, y_0 and outputs z_3, z_2, z_1, z_0 such that

$$Z = Y \times X$$

where Z, Y, X are the positive integers represented by $z_3z_2z_1z_0$, y_1y_0 and x_1x_0 using the obvious representation.

Find a minimal sum of products expression for each of z_3, z_2, z_1 and z_0. [10 marks]

Comment on the complexity of building an eight-bit multiplier using a minimal sum of products form. [3 marks]

Describe another way of building an eight-bit multiplier. [5 marks]

2 Foundations of Programming

The following is the start of a Java program which is intended to print out all prime numbers below 600.

```
public class Primes
 { private static final int SIZE=600, SQRTSIZE=25;

   public static void main(String[] args)
    { boolean[] primes = new boolean[SIZE];
      for (int i=2; i<SIZE; i++)
         primes[i] = true;

      int next = 2;
      while (next < SQRTSIZE)
       { for (int i = next; i<SIZE/next; i++)
            primes[i*next] = false;
         do
          { next++;
          } while (!primes[next]);
       }
      .
      .
```

Explain how the program is intended to work. [5 marks]

Unfortunately one value determined by the program is 589 which is not prime. Where is the bug in the program and how may it be fixed? [4 marks]

All other values determined by the program *are* prime. If the bug were left unfixed and the `final` values were adjusted so that the program attempted to find all primes below 150, would any non-prime be determined? If so, what is it? If not, why not? [4 marks]

Supply the missing part of the program so that the values which are printed out are arranged 10 numbers to a line with each number in a field of 5 places. It is not necessary to write out the part of the program which is shown above (with or without the bug). [7 marks]

3 Further Java

Describe the model for handling graphical output and interactive input in the Abstract Windowing Toolkit (AWT) for Java. Your answer should cover the use of

- hierarchies of classes
- overriding methods
- interfaces
- inner classes
- spatial hierarchy

[4 marks each]

4 Compiler Construction

Consider the grammar

```
S -> E <eof>
E -> T + E
E -> T
T -> x
```

where `S` is the starting symbol, `<eof>` is a special token marking end of input and `x` is a terminal.

Explain and find the left, right and follow sets for all non-terminals in the grammar. [5 marks]

Suppose that an SLR parser for this grammar is required. One stage on the way to constructing the parsing tables is to create the *characteristic finite state machine* (sometimes known as the LR(0) states). Do this, explaining your working clearly. You do not need to complete the SLR parsing tables. [10 marks]

Now, assuming that the parsing tables have been constructed, show what values will be placed on a stack and comment about internal state while an SLR parser using this grammar processes the input text `x+x+x<eof>`. [5 marks]

5 Data Structures and Algorithms

Describe in detail both Prim's and Kruskal's algorithms for finding a minimum cost spanning tree of an undirected graph with edges labelled with positive costs, and explain why they are correct. [7 marks each]

Compare the relative merits of the two algorithms. [6 marks]

6 Operating System Foundations

Discuss approaches to process scheduling for the following systems.

(*a*) A hard real-time system where all processes are defined statically and are periodic with known work-time per period. [4 marks]

(*b*) A shared system which runs applications controlled interactively by users. [8 marks]

(*c*) A network-attached, multimedia workstation. [8 marks]

7 Operating System Functions

Why is it important for an operating system to schedule disc requests? [4 marks]

Briefly describe each of the SSTF, SCAN and C-SCAN disc scheduling algorithms. Which problem with SSTF does SCAN seek to overcome? Which problem with SCAN does C-SCAN seek to overcome? [5 marks]

Consider a Winchester-style hard disc with 100 cylinders, 4 double-sided platters and 25 sectors per track. The following is the (time-ordered) sequence of requests for disc sectors:

{ 3518, 1846, 8924, 6672, 1590, 4126, 107, 9750, 158, 6621, 446, 11 }

The disc arm is currently at cylinder 10, moving towards 100. For each of SSTF, SCAN and C-SCAN, give the order in which the above requests would be serviced. [3 marks]

Which factors do the above disc arm scheduling algorithms ignore? How could these be taken into account? [4 marks]

Discuss ways in which an operating system can construct logical volumes which are (*a*) more reliable and (*b*) higher performance than the underlying hardware. [4 marks]

8 Mathematics for Computation Theory

Let E, F be events over a finite alphabet S. Define the events $E + F$, EF and E^*. [3 marks]

Show that:

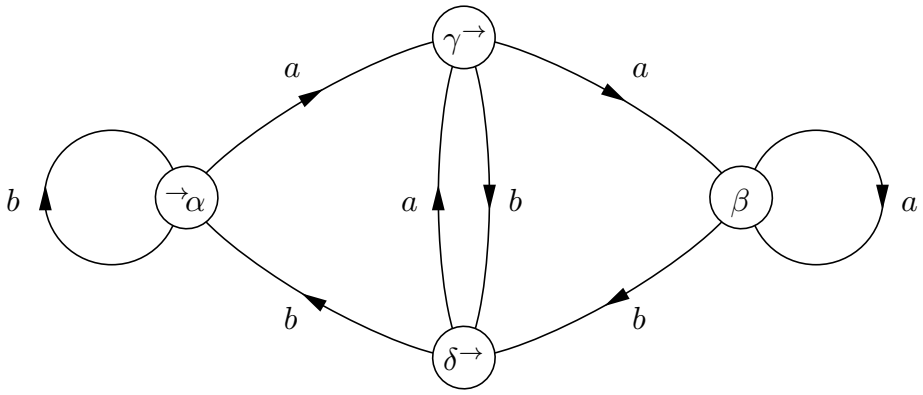
(a) $E^* = 1 + EE^*$

(b) $E(FE)^* = (EF)^*E$

[4 marks]

State Kleene's Theorem on the structure of events accepted by some Deterministic Finite Automaton (DFA). [1 mark]

Consider the following DFA:



Here α is the initial state, γ and δ the two accepting states. Show that the event accepted is

$$b^*a(a^*b\,b^*a)^*\{1 + a^*b\}.$$

[12 marks]

[Hint. If $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ is a partitioning of the transition matrix of a DFA so that A and D are square, then

$$M^* = \begin{pmatrix} (A + BD^*C)^* & A^*B(D + CA^*B)^* \\ D^*C(A + BD^*C)^* & (D + CA^*B)^* \end{pmatrix}$$

with the same partitioning. Partition the states in the order $\{\alpha, \beta\}, \{\gamma, \delta\}$. You need calculate only the upper right component of M^*.]

9 Computation Theory

Let $\mathbb{N}$ be the natural numbers $\{0, 1, 2 \ldots\}$.

What is meant by each of the following statements?

- The subset $S \subseteq \mathbb{N}$ is *recursive*.
- The subset $S \subseteq \mathbb{N}$ is *recursively enumerable*.

[5 marks]

How would you extend the definition of *recursive enumeration* to sets of computable functions? [3 marks]

A sequence of natural numbers is a total function $s : \mathbb{N} \to \mathbb{N}$. The sequence is *recursive* if and only if s is computable.

A finite sequence σ of natural numbers is specified by a pair (l, x), where $l \in \mathbb{N}$ is the number of elements, and $x : [1, l] \to \mathbb{N}$ is a function that defines those elements. The case $l = 0$ defines the null sequence.

In each of the following cases, establish whether the set defined is recursively enumerable:

(a) the set of all recursive subsets of $\mathbb{N}$ [5 marks]

(b) the set of all recursive sequences of natural numbers [2 marks]

(c) the set of all finite sequences of natural numbers [5 marks]

10 Numerical Analysis I

Define the *absolute error* ε_x and *relative error* δ_x in representing a number x. How are these errors related? Which type of error is associated with the term *loss of significance*? Define *machine epsilon* ε_m. [4 marks]

Write down expressions for the worst case errors $|\delta_{xy}|$ and $|\varepsilon_{xy}|$ in computing xy. Express your answers in terms of δ_x and δ_y. Hence find expressions for $|\varepsilon_{xy+z}|$ and $|\delta_{xy+z}|$ in terms of δ_x, δ_y and δ_z. [6 marks]

Assuming $|\,\delta_x\,| = |\,\delta_y\,| = |\,\delta_z\,| = \varepsilon_m$, find an expression for $|\,\delta_{xy+z}\,|/\varepsilon_m$. When would you expect loss of significance in computing $xy + z$? [3 marks]

The formula

$$\tilde{f}'(x) = \frac{f(x+h) - f(x)}{h}$$

is used to compute $f'(x)$ for a certain function $f(x)$ when $\varepsilon_m = 10^{-10}$. The formula is applied at $x = 0.2$ where $f(0.2) = 1.1$.

(a) If $h = 10^{-3}$ then $\tilde{f}'(0.2) \simeq 8.44$.

(b) If $h = 10^{-8}$ then $\tilde{f}'(0.2) \simeq 8.40$.

However, it is known that $f'(0.2) = 8.42$ to 3 significant digits. Why is (a) such a poor estimate? Why is (b) such a poor estimate? [4 marks]

Suggest a more suitable value for h. Roughly how many correct significant decimal digits would you expect to get in your answer? [3 marks]

11 Computer Graphics and Image Processing

Describe an algorithm for performing scan conversion of a set of 3D polygons, including details of clipping, projection, and the underlying 2D polygon scan conversion algorithm. You may assume that you are given the colour of each polygon and that no lighting calculations are required. *Ray tracing* is not an acceptable answer to this question. [20 marks]

12 Software Engineering II

Consider this program over integer variables:

```
k := K;
x := X;
z := 1;
while k <> 0 do
  begin
    k := k-1;
    z := z*x
  end
```

Given that the loop invariant is $z \times x^k = X^K$, show that executing this program stores the value of X^K in the variable z. [5 marks]

It is proposed to insert the following code just before the assignment `k := k-1`:

```
while even(k) do
  begin
    k := k/2;
    x := x*x
  end
```

State the loop invariant of this inner loop and show that the modified program still stores the value of X^K in z. [7 marks]

Briefly describe formal specification languages, top–down design and fault avoidance techniques, indicating their respective roles in a software development project. [8 marks]

13 Natural Language Processing

The following context-free grammar and lexicon generates the examples below it with multiple derivations and therefore multiple associated interpretations.

(a) S $\rightarrow$ NP VP

(b) NP $\rightarrow$ Det N

(c) NP $\rightarrow$ NP PP

(d) N $\rightarrow$ N N

(e) VP $\rightarrow$ V NP

(f) VP $\rightarrow$ V NP PP

(g) VP $\rightarrow$ VP PP

(h) PP $\rightarrow$ P NP

N $\rightarrow$ car | park | tree | boy | toy | morning | ...
V $\rightarrow$ hit | ...
P $\rightarrow$ in | with | ...
Det $\rightarrow$ a | the | ...

a car hit the tree in the park
the boy hit the toy car park with a toy car in the morning

Describe how a probabilistic version of the context-free grammar (PCFG) can be created, defining the constraints which must hold for the resulting PCFG to be interpretable as a stochastic language model. [8 marks]

How accurate would the resulting PCFG be at assigning the semantically appropriate derivations the highest probability for the examples and other structurally similar sentences? [6 marks]

Define an improved probabilistic model for discriminating alternative derivations. What problems would arise in the implementation of this model? [6 marks]

END OF PAPER