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

Tuesday 1 June 1999 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 Software Engineering I

State the advantages of the waterfall model. [10 marks]

State the disadvantages of the waterfall model. [4 marks]

What is the main criterion for deciding whether or not to use the waterfall model in a software development project? [3 marks]

Explain whether there would be any difference for a hardware development project. [3 marks]

2 Foundations of Programming

The Java program below has been written by someone new to list-processing. This test program is intended to set up a list whose elements are 4, 7 and 11.

```
public class ListTest
 { public static void main(String[] args)
    { Link start = null;
      start.put(4);  start.put(7);  start.put(11);
      System.out.println("List elements:  " + start);
    }
 }

class Link
 { private int val;
   private Link next;

   public Link(int n)
    { this.val = n;
      this.next = null;
    }

   public void put(int k)
    { if (this == null)
         this = new Link(k);       // Error noted here
      else
         this.next.put(k);
    }

   public String toString()
    { return (this == null ? "" : "  " + this.next.toString());
    }
 }
```

The Java compiler gives a single error message, complaining about the statement indicated. What is the problem? Explain why there is more to fixing the program than merely changing this statement. [5 marks]

Making the minimum number of changes, modify the program so that it works approximately in the way the author intended. [6 marks]

Provide for class `Link` a method `sum()` which returns the sum of the elements in the list. [3 marks]

Provide for class `Link` a method `reverse()` which returns a new list whose elements are in the reverse order of those in the original list. [6 marks]

3 Further Java

Describe the facilities in Java for restricting concurrent access to critical regions. Explain how shared data can be protected through the use of objects. [8 marks]

The built-in facilities for restricting concurrency in Java allow only one thread at a time to be executing within the critical region. A different approach is to distinguish *shared* and *exclusive* access to a critical region: any number of *readers* may share access at the same time, but only one *writer* may acquire exclusive access (excluding any readers while it does so).

Specify a `MultiSync` class in Java with methods to acquire and release both read and write access, and sketch its implementation. [6 marks]

Derive a `MultiBuffer` class that extends `MultiSync` with methods to store and read a data field, ensuring that any locks are released when a waiting thread is interrupted. Your example may use a simple data field such as an integer but you should explain why such an elaborate scheme of concurrency control is unnecessary in this case. [6 marks]

4 Compiler Construction

It is commonly suggested that Algol-60 call-by-name can be modelled by passing a function as a call-by-value parameter. Show how a program containing a definition

```
int f(int x:name) { ... x ... x ... }
```

of `f` (where `x` occurs only in Rvalue context) and a call `f(`e`)` to `f` can be replaced by an equivalent definition and call using only call-by-value. [6 marks]

Most such explanations assume that the uses of `x` within `f` occur only in Rvalue context. However, Algol-60 also permits the equivalent of

```
int g(int x:name) { if (p) { ... x := x+1; x := -x; ... }
                          return x;
                    }
```

and calls like `g(a[k()])` which, when `p` is `true`, would have the effect of calling `k()` five times and consequent access to five (possibly different) subscripts of array `a[]`. Develop your explanation for the first part of this question to cover also the case of a call-by-name parameter being used in both Lvalue and Rvalue contexts. [Hint: note that when `p` is `false` then the actual parameter to `g` need not be an Lvalue, so you may need two parameterless procedure arguments ("thunks").] [8 marks]

Using the previous part or otherwise, give a translation of a definition and call `h(`e`)` using call-by-value-result (Ada `in out` mode) with no uses of the address-of (`&`) operator other than those involved in call-by-name. Your explanation is allowed to deviate from call-by-value-result by allowing side-effects in e to take place twice. [6 marks]

5 Data Structures and Algorithms

Outline the mechanism used in the Burrows–Wheeler block compression algorithm, illustrating your description by applying it to the string ALFALFA. [14 marks]

Briefly discuss the advantages and disadvantages of the Burrows–Wheeler algorithm compared with other commonly used compression methods. [6 marks]

6 Operating System Foundations

Producer and consumer processes interact via an N-slot cyclic buffer. Semaphores are defined and initialised as follows:

$$\begin{aligned} &lock : semaphore &:= 1 \\ &spaces : semaphore &:= N \\ &items : semaphore &:= 0 \end{aligned}$$

For the following programs indicate where mutual exclusion and condition synchronisation are being attempted and explain how the system may fail.

producer code	consumer code
produce item	WAIT (*lock*)
WAIT (*lock*)	WAIT (*items*)
WAIT (*spaces*)	remove item
insert item	SIGNAL (*spaces*)
SIGNAL (*items*)	SIGNAL (*lock*)
SIGNAL (*lock*)	consume item

[8 marks]

Write a monitor to manage the N-slot buffer. Discuss why the problems you pointed out in the previous part do not arise in the monitor implementation. [12 marks]

7 Operating System Functions

The following are three ways which a file system may use to determine which disk blocks make up a given file.

(*a*) chaining in a map

(*b*) tables of pointers

(*c*) extent lists

Briefly describe how each scheme works. [3 marks each]

Describe the benefits and drawbacks of using scheme (*c*). [6 marks]

You are part of a team designing a distributed filing system which replicates files for performance and fault-tolerance reasons. It is required that rights to a given file can be revoked within T milliseconds ($T \geqslant 0$). Describe how you would achieve this, commenting on how the value of T would influence your decision. [5 marks]

8 Mathematics for Computation Theory

Let S be a finite alphabet. Define

(a) the set of events E over S

(b) acceptance of an event E by a deterministic finite automaton (DFA) M

(c) the regular operators on events

(d) the set of regular events over S

[9 marks]

State Kleene's Theorem. [2 marks]

Suppose that the event E is accepted by an N-state DFA $M \equiv (Q, S, \iota, f, A)$. Show that if E is non-empty, then M must accept some word w such that $\ell(w) < N$.

[5 marks]

Let regular events E, E' over the same alphabet S be accepted by DFA M, M' respectively. Show that it is decidable whether $E = E'$. [4 marks]

[If you use the Pumping Lemma it should be clearly stated.]

9 Computation Theory

Define the *primitive recursive* and *partial* ($\mu-$) *recursive* functions. [6 marks]

Suppose you are given a Turing machine with state set Q and k-symbol alphabet S whose action is defined by transition functions

$$\begin{aligned} q' &= f(q,s) \in Q \uplus \{H\} && \text{(disjoint union)} \\ s' &= r(q,s) \in S && \text{(replacement symbol)} \\ d' &= d(q,s) \in \{L,R,C\} && \text{(movement)} \end{aligned}$$

where the head moves to L or R on the tape unless $q' = H$, in which case $d' = C$ and the machine stops.

Extend the action of the machine by an additional state symbol D so that for all $s \in S$,

$$\begin{aligned} f(H,s) &= f(D,s) = D \\ r(H,s) &= r(D,s) = s \\ d(H,s) &= d(D,s) = C \end{aligned}$$

Show that the action of the Turing machine as extended in this way can be described by a primitive recursive function $T(t,x)$, where t is a step counter and x is a code specifying the initial configuration. [10 marks]

Hence show that computation by any Turing machine may be represented by a partial recursive function. [4 marks]

10 Natural Language Processing

Write about *two* of the following. Describe how each technique can be used in the solution of a natural language processing (sub)task and the problems which remain.

(*a*) Active chart parsing

(*b*) Part-of-speech disambiguation using finite-state machines

(*c*) Probabilistic context-free grammar

(*d*) Plan recognition for discourse comprehension

[10 marks each]

11 Computer Graphics and Image Processing

Give the formula for a Bezier cubic curve.

Derive the conditions necessary for two Bezier cubic curves to join with (*a*) just $C0$-continuity and (*b*) $C1$-continuity.

Give a geometric interpretation of each condition in terms of the locations of control points.

Explain (mathematically) why a Bezier cubic curve is guaranteed to lie within the convex hull of its control points. [8 marks]

Basic ray tracing uses a single sample per pixel. Describe *four* distinct reasons why one might use multiple samples per pixel. Explain the effect that each is trying to achieve, and outline the mechanism by which it achieves the effect. [8 marks]

Describe the differences in the computational complexity of the depth sort and binary space partition (BSP) tree algorithms for polygon scan conversion.

If you were forced to choose between the two algorithms for a particular application, what factors would be important in your choice? [4 marks]

12 Digital Electronics

An *arbiter* is a functional unit that allows requesters to lock resources while they are in use.

(*a*) Describe the inputs and outputs of an arbiter for a single resource which has two requesters. [5 marks]

(*b*) Describe *two* policies that the arbiter might implement when both requesters request the resource simultaneously. One of these should require some internal state. [5 marks]

(*c*) Give a state diagram for a policy which requires internal state assuming synchronous operation. [5 marks]

(*d*) Choose *one* of the policies in (*b*) and derive equations for outputs and next state control using any type of storage element you choose. [5 marks]

13 Numerical Analysis I

The mid-point rule can be expressed in the form

$$I_n = \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} f(x)dx = f(n) + e_n$$

where

$$e_n = f''(\theta_n)/24$$

for some θ_n in the interval $(n - \frac{1}{2}, n + \frac{1}{2})$. Assuming that a formula for $\int f(x)dx$ is known, and using the notation

$$S_{p,q} = \sum_{n=p}^{q} f(n),$$

describe a method for estimating the sum of a slowly convergent series $S_{1,\infty}$, by summing only the first N terms and estimating the remainder by integration.

[6 marks]

Assuming that $f''(x)$ is a positive decreasing function, derive an estimate of the error $|E_N|$ in the method. [5 marks]

Given

$$\int \frac{dx}{x(x+2)} = -\frac{1}{2}\log_e(1 + \frac{2}{x})$$

illustrate the method by applying it to

$$\sum_{n=1}^{\infty} \frac{1}{n(n+2)}.$$

Verify that $f''(x)$ is positive decreasing for large x, and estimate the integral remainder to be added to $S_{1,N}$. [You may assume $\log_e(1 + \lambda) \simeq \lambda$ for λ small.]

[6 marks]

To 2 significant digits, how large should N be to achieve an absolute error of approximately 1.8×10^{-11}? [3 marks]