COMPUTER SCIENCE TRIPOS Part Ib

Thursday 3 June 1999 1.30 to 4.30

Paper 6

Answer **five** *questions.*
No more than **two** *questions from any one section are to be answered.*
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.*

SECTION A

1 Data Structures and Algorithms

Splay trees provide an adaptive mechanism for the representation of sets of character strings. Outline the structure used by splay trees and briefly describe how the *insert*, *lookup* and *delete* operations work. [10 marks]

Discuss the advantages and disadvantages of using splay trees compared with the use of hash tables for the representation of sets of distinct character strings. [10 marks]

2 Computer Design

What is the difference between serial and parallel data transmission? [4 marks]

RS232 is a serial communications protocol. Explain how the letter "A" (ASCII code 65 in decimal) is transmitted if one start bit and one stop bit are to be used. [5 marks]

Parallel data transmission is typically used to communicate with main memory (DRAM). Explain the principal operations involved in reading a burst of consecutive data words from DRAM. [6 marks]

What is a *cache* and why would it typically wish to read a block of data rather than a single word from the DRAM? [5 marks]

3 Digital Communication I

What is a *hierarchical address space*? Give an example of an address space which is hierarchical and one which is not. [3 marks]

What is the Address Resolution Protocol? Describe its operation when used to resolve IP addresses to Ethernet addresses. Pay particular attention to the freshness of information. [7 marks]

Information is transferred via a long, error-prone communication link. The link has a data rate of 10 Mbps and a constant delay. The bit error rate on the link is 1 bit in 10^4. A forward error correcting coder is available which can act in the following settings:

Data rate	Error rate	Code rate
10 Mbps	10^{-4}	unity (no coding)
5 Mbps	10^{-5}	half

A simple ARQ protocol is used over the link. Packets have 32-bit CRCs. You may assume that the undetected error rate is less than 1 in 10^{20}, that is, effectively zero.

Information is sent in 1000-bit packets, with a window of one packet. At what link delay would it be beneficial to use the FEC coder? [10 marks]

4 Computer Graphics and Image Processing

You have a cone of height one unit; apex at the origin; and base of diameter one unit centred on the negative z-axis.

You wish to transform this cone so that the apex is at $(-1, 3, 7)$, the base is centred at $(2, 7, -5)$, and the base's radius is four units. What transformations are required to achieve this and in what order should they be performed? [8 marks]

Describe an algorithm which clips an arbitrary polygon against an arbitrary *convex* polygon (in 2D). [8 marks]

Will your algorithm correctly clip an arbitrary polygon against an arbitrary *non-convex* polygon? If so, explain why and demonstrate an example which illustrates that it does work in such cases. If not, explain why not and outline how your algorithm could be extended to perform clipping in such cases. [4 marks]

SECTION B

5 Comparative Programming Languages

Early programming languages had relatively poor facilities for type checking, data abstraction, data hiding and encapsulation. Explain the meaning of these terms and discuss their importance. [6 marks]

Outline the key features that a language must have to be called object-oriented and briefly discuss to what extent C++, Java and Smalltalk have them. [7 marks]

Briefly discuss some of the reasons why C++ programs typically run faster than equivalent programs written in Java or Smalltalk. [7 marks]

6 Compiler Construction

It is desired to obtain an unambiguous context-free grammar G' which generates the same strings as the following grammar G with start symbol `S`.

```
S -> E
E -> ( E ) | [ E ] | E * E | a | b | c
( E ) -> ( + E )
[ E ] -> [ - E ]
```

Define a suitable G' or explain why G already satisfies the criterion. [6 marks]

Write a context-free (Type 2) grammar which describes floating-point numbers of the form $[\pm]dd^*[.d^*][e[\pm]dd^*]$ where d stands for decimal digit and d^* stands for zero or more decimal digits. $[\cdots]$ means that the enclosed item is optionally present in the floating-point number. [7 marks]

Sketch a recursive descent parser for the following grammar H with start symbol `S`. You should assume the existence of a routine `lex()` which sets variable `token` to one of '`1`', '`2`', '`(`', '`)`', '`-`' or `eof`.

```
P -> 1 | 2 | (E)
E -> P | E - P
S -> E eof
```

[7 marks]

7 Prolog for Artificial Intelligence

A trinary tree is constructed from 3-ary compound terms $\mathsf{n}(a, b, c)$ called nodes, where components a, b and c are either nodes or integers. Assume that integer components are restricted to the values 0 and 1.

Write a Prolog program to return a list of all the 0's and a list of all the 1's in a given tree. For example, the goal `enum(n(n(0, 1, 0), 1, 0), X, Y)` should instantiate `X` to `[0, 0, 0]` and `Y` to `[1, 1]`. The program should use difference lists. [10 marks]

A terminal node of the trinary tree is said to be of *odd parity* if the number of its 1 components is an odd number. For example, `n(1, 1, 1)` is of odd parity, and `n(1, 0, 1)` is not of odd parity. Write a Prolog program to count the number of terminal nodes in a tree that have odd parity. For example, the goal `odd(n(n(0, 1, 0), 1, 0), X)` should instantiate `X` to 1. [10 marks]

8 Databases

Give *three* ways in which an Object Data Model can extend the semantics offered by the 1992 SQL Standard. [5 marks]

Compare and contrast the facilities of the ODMG Database Standard and of the proposed SQL3 standard for representing the data and behaviour of objects. Illustrate your answer by considering the employee records of a transport business. [12 marks]

What drawbacks might be associated with the use of an OODB for this data? [3 marks]

SECTION C

9 Semantics of Programming Languages

Give the rules for *call-by-value* and *call-by-name* evaluation of function applications in the language LFP combining state-manipulating commands with function abstraction and application. [5 marks]

Is the call-by-value evaluation relation contained in the call-by-name evaluation relation? Is the call-by-name evaluation relation contained in the call-by-value evaluation relation? Justify your answer in each case. [6 marks]

Briefly describe how a type system for LFP can be used to detect statically that some LFP terms do not cause side-effects on the state under call-by-name evaluation. What happens for call-by-value evaluation? [9 marks]

10 Logic and Proof

Describe the role of Herbrand models in mechanical theorem proving. What may we infer when a set of clauses has no Herbrand model? [3 marks]

The remainder of this question concerns using clause methods to determine whether or not the formula

$$\exists x\,[P(x) \wedge Q(x)] \rightarrow \exists x\,[P(f(x,x)) \vee \forall y\,Q(y)]$$

is a theorem.

Convert the problem into clause form. Justify each step you take and explain in what respect the set of clauses is equivalent to the original problem. [4 marks]

Describe the Herbrand universe for your clauses. [3 marks]

Produce a resolution proof from your clauses, or give reasons why none exists. [4 marks]

Exhibit a Herbrand model for your clauses, or give reasons why none exists. [6 marks]

11 Complexity Theory

Explain what is meant by a deterministic and a non-deterministic Turing Machine and the idea of such machines solving a decision problem. [7 marks]

If a non-deterministic Turing Machine solves a certain problem in at most N time-steps, what information must be noted to document the exact state of the machine at each stage as it performs the calculation? [5 marks]

Part of the information you have just identified will be the sequence of states $q_0, q_1, \ldots$ that the machine goes through. Taking account of the fact that the machine is non-deterministic show how

(a) this part of the information can be represented by the values of a number of boolean variables, and

(b) a formula in the style used in the problem 3-SAT can be written down to ensure that the sequence of states is one that does correspond to a valid computation of the machine.

[*Marks will be deducted if you attempt to document how to extend your demonstration to cover other aspects of the machine's behaviour.*] [8 marks]

12 Foundations of Functional Programming

(*a*) Give λ-terms Y, K, T and I satisfying the following equalities for all terms M and N:

$$\begin{aligned} YM &= M(YM) \\ KMN &= M \\ TMN &= NM \\ IM &= M \end{aligned}$$

[4 marks]

(*b*) A λ-term is *defined* if it has a head normal form. For each of the following terms, state whether or not it is defined, giving justification for your answer.

$$Y \qquad YK \qquad YT \qquad YI$$

[8 marks]

(*c*) A λ-term M is *solvable* if there exist variables $x_1, \ldots, x_m$ and terms $N_1, \ldots, N_n$ such that

$$(\lambda x_1, \ldots, x_m.M)N_1 \ldots N_n = I$$

For those terms in (*b*) that are solvable, exhibit the variables and terms that establish this. For those that are not, explain why they are not solvable.

[8 marks]