COMPUTER SCIENCE TRIPOS Part Ib

Wednesday 8 June 2005 1.30 to 4.30

Paper 5

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.*

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

STATIONERY REQUIREMENTS
Script Paper
Blue Coversheets
Tags

SECTION A

1 Data Structures and Algorithms

(*a*) Explain how a Boolean matrix can be used to represent the edges of a finite directed graph whose vertices are numbered 1 to n. [2 marks]

(*b*) Describe Warshall's algorithm to convert the matrix representing a graph to one that represents its transitive closure, and carefully explain why the algorithm works. [6 marks]

(*c*) Outline Floyd's algorithm, without proof of correctness, to find the cost of the cheapest path between any two vertices of a directed graph where the edges carry non-negative costs. [4 marks]

(*d*) It is required to construct a matrix R that encodes a path with the minimum number of edges from any vertex i to any other vertex j. R_{ij} will be zero if no path exists from vertex i to vertex j; otherwise, R_{ij} will hold the vertex number of the next vertex of a minimal path from i to j. Suggest an algorithm to compute R from a given Boolean matrix M. [8 marks]

2 Computer Design

(*a*) What is the difference between a *control hazard* and a *data hazard*? [4 marks]

(*b*) How are data and control hazards handled for the following two processors with their respective pipelines?

The N-105 processor pipeline:

instruction fetch	register fetch, decode, execute memory access and write back

The ARM9 processor pipeline:

instruction fetch	decode	execute	memory access	write back

[8 marks]

(*c*) If a load instruction causes a cache miss, what impact does it have on the pipeline? [3 marks]

(*d*) What is the structure of a TLB (Translation Lookaside Buffer)? [2 marks]

(*e*) What impact does a TLB miss have on the pipeline? [3 marks]

3 Digital Communication I

It is proposed to send information across a fixed delay channel using a simple (window of 1) ARQ protocol with a transmitter timeout of T. That is, if the transmitter does not receive an acknowledgement for a packet within time T of sending the packet, it retransmits.

The delay of the underlying channel is τ, the data rate is B and the packet size is p bits. Bit errors in the channel are independent and packets of size p have a packet error rate of e. Errors in the small acknowledgement packets are rare enough to be discounted in this analysis.

(a) What is the expected throughput of the ARQ protocol if e is zero? [4 marks]

(b) What is the expected throughput if e is non-zero, but small enough that e^2 is negligibly small? [4 marks]

(c) How could a forward error code help the throughput of the ARQ scheme? [2 marks]

(d) What is meant by the term *code rate* of a forward error code? [2 marks]

(e) What code rate must a code which squared the error rate have in order to improve throughput of the ARQ scheme? [4 marks]

(f) If the forward error coder adds delay, how will this affect performance? [4 marks]

4 Concurrent Systems and Applications

(*a*) Mutual exclusion is an important consideration in many multi-threaded processes.

(*i*) Describe the syntax and semantics of each of the different ways of using the `synchronized` keyword in Java. [5 marks]

(*ii*) What is a *re-entrant* mutual exclusion lock and why is it helpful that the locks used by the Java Virtual Machine to implement synchronized methods be re-entrant? [2 marks]

(*iii*) Accesses to fields of most data types in Java are atomic but some are not. Give an example of a field access that is not atomic and explain how read and write access can be achieved in a thread-safe fashion. [2 marks]

(*iv*) Recent editions of the Java language include *generics*. What is the scope of the mutual exclusion caused by the use of the `synchronized` keyword in a generic class definition? [1 mark]

(*b*) *Deadlock* can occur in multi-threaded applications.

(*i*) What *four* properties hold when deadlock exists? [4 marks]

(*ii*) Which of these are properties of the Java language and which depend on the program being executed? [1 mark]

(*iii*) A practical strategy to avoid deadlock is to enforce an ordering on acquiring locks. Explain how this ensures that deadlock is never possible. [2 marks]

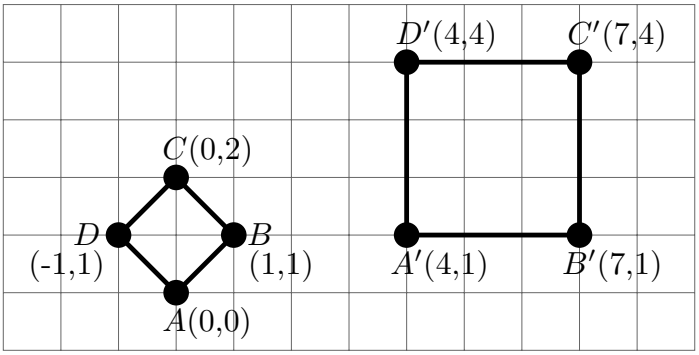
(*c*) Mutual exclusion locks are language-level features. Explain how they can be implemented in terms of **either** Counting Semaphores provided by the operating system **or** atomic compare-and-swap operations provided by the hardware. [3 marks]

SECTION B

5 Computer Graphics and Image Processing

(*a*) Describe, in detail, an algorithm to clip a straight line against an axis-aligned rectangle. [10 marks]

(*b*) Explain why homogeneous coordinates are used for handling geometric transformations. [3 marks]



(*c*) Give a matrix, or a product of matrices, which will transform the square $ABCD$ to the square $A'B'C'D'$. [4 marks]

(*d*) Show what happens if the same transformation is applied to the square $A'B'C'D'$. [3 marks]

6 Compiler Construction

(*a*) Explain how a parse tree representing an expression can (*i*) be converted into stack-oriented intermediate code and then (*ii*) be translated into simple machine code for a register-oriented architecture (e.g. ARM or IA32) on an instruction-by-instruction basis. Also indicate how this code might be improved to remove push–pop pairs introduced by (*ii*). Your answer need only consider expression forms encountered in the expression:

```
h(a, g(b), c) * 3 + d
```

[12 marks]

(*b*) In Java, expressions are evaluated strictly left-to-right. Consider compiling the function `f` in the following Java class definition:

```
class A
{
    static int a,b;
    void f() { ... <<C>> ... }
    int g(int x) { ... a++; ... }
};
```

Indicate what *both* the intermediate code *and* (improved as above) target code might be for `<<C>>` for the cases where `<<C>>` is:

(*i*) `b = g(7) + a;`

(*ii*) `b = a + g(7);`

(*iii*) `b = (-g(7)) + a;`

(*iv*) `b = a - g(7);`

Comment on any inherent differences in efficiency at both the intermediate code and target code levels.

[8 marks]

7 Comparative Programming Languages

Most large programs that have been written with considerable care and thoroughly checked still seem to contain bugs at a rate of over one per 3000 lines of source code. Systems involving hundreds of millions of lines of code can thus be expected to contain tens of thousands of potentially catastrophic errors.

(*a*) List several kinds of programming errors that can appear in programs and discuss their relative importance in relation to the long-term reliability of a large application program. [7 marks]

(*b*) Suggest potential ways by which programmers may reduce the number of programming errors they make, paying particular attention to language features that might help, extra features in program development systems and possible changes in overall system architecture. [8 marks]

(*c*) In what ways would you expect languages 25 years from now to differ from those that are currently popular? [5 marks]

8 Databases

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

(*b*) Describe *two* differences and *two* similarities between the relational algebra and SQL. [4 marks]

(*c*) Suppose that $S(a,\ b,\ \ldots)$ and $R(a,\ \ldots)$ are relations (the notation indicates that attribute a is in the schema of both S and R, while attribute b is only in the schema of S). Suppose that v is a value; is the following equation always valid?

$$\sigma_{(a=v \text{ or } b=v)}(R \bowtie S) = (\sigma_{a=v}(R)) \bowtie (\sigma_{b=v}(S))$$

If yes, provide a short proof. If no, provide a counter-example. [2 marks]

(*d*) Various *normal forms* are important in relational schema design.

(*i*) Define Third Normal Form (3NF). [3 marks]

(*ii*) Define Boyce-Codd Normal Form (BCNF). [3 marks]

(*iii*) For databases with many concurrent update transactions, explain why schemas in normal form are important for good performance. [3 marks]

SECTION C

9 Logic and Proof

(*a*) In order to prove the following formula by resolution, what set of clauses should be submitted to the prover? Justify your answer briefly.

$$\forall x\,[P(x) \vee Q \rightarrow \neg R(x)] \wedge \forall x\,[(Q \rightarrow \neg S(x)) \rightarrow (P(x) \wedge R(x))] \rightarrow \forall x\,S(x)$$

[7 marks]

(*b*) Derive the empty clause using resolution with the following set of clauses, or give convincing reasons why it cannot be derived.

$$\{\neg P(x,x)\} \quad \{P(x,f(x))\} \quad \{\neg P(x,y), \neg P(y,z), P(x,z)\}$$

[6 marks]

(*c*) Derive the empty clause using resolution with the following set of clauses, or give convincing reasons why it cannot be derived. (Note that a and b are constants.)

$$\{\neg P(a)\} \quad \{Q(a)\} \quad \{R(b)\} \quad \{S(b)\}$$
$$\{\neg Q(x), P(x), \neg R(y), \neg Q(y)\} \quad \{\neg S(x), \neg R(x), Q(x)\}$$

[7 marks]

10 Complexity Theory

(*a*) Explain what it means to say that a problem is

(*i*) NP [2 marks]

(*ii*) NP-Complete [2 marks]

(*b*) Define the standard problem 3-SAT and describe how you would take an instance of it and derive an integer n that you would use in any formulae relating to the cost of solving that instance. [3 marks]

(*c*) What is a non-deterministic Turing Machine? Supposing that some computation of such a machine takes N steps, what information needs to be reported to describe exactly how the computation proceeded? In what way is this relevant to the problem of solving an *arbitrary* NP problem? [7 marks]

(*d*) Sketch a proof of Cook's result, that the problem 3-SAT is NP complete. Justify that any transformations you introduce are polynomial. [6 marks]

11 Semantics of Programming Languages

(*a*) Give a call-by-name small-step operational semantics and typing rules for the syntax below.

$$e \quad ::= \quad \textsf{fn}\ x \Rightarrow e \mid e_1\ e_2 \mid x$$

[5 marks]

(*b*) State and prove Determinacy. [6 marks]

(*c*) State and prove Type Preservation. [9 marks]

Any standard lemmas you use must be stated precisely, but need not be proved.

12 Foundations of Functional Programming

(*a*) Give a lambda expression that can be used to form the composition of two functions. [1 mark]

(*b*) Suppose that the lambda expression you have given above can be referred to using the name `B`. One way of representing the natural numbers as lambda expressions involves for instance having the number "3" represented by a term $\lambda f.\mathtt{B}f(\mathtt{B}ff)$ so that a numeral when applied to an argument f composes f with itself the given number of times.

In this scheme, write out lambda expressions that will serve as 0, 1 and 2. [3 marks]

(*c*) Present and explain lambda expressions that find the successor to a number represented as in part (*b*) and that add two numbers together. [6 marks]

(*d*) If m and n are two lambda expressions that both represent numbers in this style, what interpretation can be placed on the term $(m\ n)$? Explain and justify your claim. [4 marks]

(*e*) Explain how it is possible to produce a lambda expression that, given the representation of a non-zero number k, produces an expression that behaves like $k - 1$. [6 marks]

END OF PAPER