COMPUTER SCIENCE TRIPOS Part II

Tuesday 2 June 2009 1.30 to 4.30

COMPUTER SCIENCE Paper 7

*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.*

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 cover sheets
Tags

SPECIAL REQUIREMENTS
Approved calculator permitted

1 Additional Topics

(*a*) Briefly describe what is meant by *active* and *passive* RFID tags. [2 marks]

(*b*) Some RFID manufacturers now produce *semi-active* RFID tags, where a battery is used to power the microelectronics but backscattering is used for all radio communications. Give *two* advantages and *two* disadvantages of such tags. [4 marks]

(*c*) Consider a typical binary tree search applied to identify all RFID tags within range of a transmitter. Each request takes the form [**REQ** | **F** | **X**], where REQ is a c-bit command ID, F is the $f < K$ filter bits and X is a $(K - f)$ bit sequence of 1s. Any response then has the form [**RESP** | **F** | **I**], where RESP is a c-bit command ID, F is the first f bits of the replying tag's ID and I represents the remaining $(K - f)$ bits of that ID.

In an attempt to increase efficiency, a manufacturer proposes that the reader just send [**REQ** | **F**] and the tags immediately respond with [**I**].

(*i*) What addition would you have to make to the communications protocol for this to work? What would its overhead be in bits? [3 marks]

(*ii*) Derive an expression for the proportional reduction in search time that this new scheme would provide. Estimate the value of the ratio for a typical tag on the market today. [5 marks]

(*d*) In a probabilistic RFID scheme, the reader transmits the number of slots in a round, N. RFID tags choose a slot uniformly at random and transmit their ID in it. Suggest how to estimate the number of tags in range based *only on one round* at the reader.

[Hint: Consider the expected number of slots with a given property such as being empty or containing a collision.]

[6 marks]

2 Advanced Graphics

(*a*) State the Jordan curve theorem. [1 mark]

(*b*) Given point V and simple convex planar polygon $P=\{v_0, v_1, \ldots, v_{n-1}\}$ in $\mathbb{R}^3$, express:

(*i*) A test for whether V is coplanar with P. [1 mark]

(*ii*) A test for whether V lies *strictly inside* P. [2 marks]

(*iii*) A test for whether V lies *on the border* of P. [1 mark]

(*c*) (*i*) Describe an algorithm for ray-tracing a complex CSG (Constructive Solid Geometry) shape. How could your algorithm be represented by a state machine? [4 marks]

(*ii*) Identify *three* Boolean operations that your algorithm would support between primitives. [1 mark]

(*iii*) Would your algorithm perform ray-primitive intersections in local, eye, screen, or world co-ordinates? Why? [2 marks]

(*d*) (*i*) Show that the closed uniform B-Spline of degree 2 and with knot vector $\{0, 0, 0, 1, 1, 1\}$ is a quadratic Bézier curve. [6 marks]

(*ii*) Sketch the basis functions of the curve's coefficient polynomials. Accuracy is not critical. [2 marks]

3 Advanced Systems Topics

(*a*) The size and rate of growth of routing tables is an issue of considerable concern for inter-domain routing in the Internet.

(*i*) Describe at least *three* factors that contribute to this problem. [3 marks]

(*ii*) How might a distinction between Locators and Identifiers help address this problem? [4 marks]

(*iii*) What are the costs and risks of implementing a Loc/ID split? [4 marks]

(*b*) BGP employs a mechanism called *Route Flap Damping* (RFD) to reduce route instabilities.

(*i*) Describe how RFD works. [4 marks]

(*ii*) Describe some problems with RFD. [2 marks]

(*iii*) Describe another approach to reducing the number of BGP updates. [3 marks]

4 Artificial Intelligence II

Evil Robot has almost completed his Evil Plan for the total destruction of the human race. He has two nasty chemicals, which he has imaginatively called A and B and which are currently stored in containers 1 and 2 respectively. All he has to do now is mix them together in container 3. His designer—an equally evil computer scientist—has equipped Evil Robot with a propositional planning system that allows him to reason about the locations of particular things and about moving a thing from one place to another.

(*a*) Explain how this problem might be represented within a propositional planning system. Give specific examples of the way in which the start state and goal can be represented. [5 marks]

(*b*) Describe in detail an algorithm that can be used to find a plan using this form of representation. [5 marks]

(*c*) Give a specific example of a *successor-state axiom* using the representation you suggested in part (*a*). [2 marks]

(*d*) Explain why in this particular planning problem it might be necessary to include one or more *precondition axioms* and give an example of such an axiom using your representation. [2 marks]

(*e*) Explain why in this particular planning problem it might be necessary to include one or more *action exclusion axioms* and give an example of such an axiom using your representation. Suggest why it might be unwise to include too many axioms of this type, and explain how a reasonable collection of such axioms might be chosen in a systematic way. [4 marks]

(*f*) Explain how in this problem it might be possible to include *state constraints* as an alternative to action exclusion axioms, and give a specific example of such a constraint using your representation. [2 marks]

5 Bioinformatics

(*a*) Discuss, with one example, the complexity of the Nussinov algorithm for RNA folding. [5 marks]

(*b*) In the context of algorithms on strings, what is the advantage of using spaced seeds in database search? [3 marks]

(*c*) Hidden Markov models (HMM) are used to identify genes in genome sequencing projects.

(*i*) Describe how you would build a hidden Markov model to identify genes in a genome sequence. [7 marks]

(*ii*) How would you assess the sensitivity and specificity performance of the HMM? [5 marks]

6 Business Studies

(*a*) Distinguish between *debt* and *equity* financing for a young company. [5 marks]

(*b*) You have won a contract to write and supply some software and set up a company to do so. The contract is worth £100,000, with 30% payable at start, 20% at a milestone expected to be completed in month 3 after starting, 40% on delivery expected in month 6 and 10% 1 month after delivery. You will need to employ two contract programmers at a rate of £2,500 each per month (plus overheads) for the duration of the contract.

(*i*) Draw up an outline cash flow budget. What is the working capital requirement? [5 marks]

(*ii*) You raise investment of £15,000 in the company and arrange a bank loan facility up to another £15,000 (ignore bank charges and loan interest for this question). You purchase £10,000 of capital equipment initially (computers etc). Draw up the balance sheet at the end of month 6. [5 marks]

(*iii*) The project unfortunately takes an additional two months before passing the milestone. Draw up a revised cash-flow budget including the funds raised and purchases made as specified in part (*b*)(*ii*). What effect does this have on the working capital requirement? What options do the Directors have if the bank refuses to extend the loan? [5 marks]

7 Comparative Architectures

(*a*) What dependencies exist between the instructions in the code fragment below? Identify both true data dependencies and name dependencies, and for each name dependence indicate whether it is an antidependence or an output dependence. [4 marks]

```
LI R1, 25         /* R1=25 */
LI R2, 8          /* R2=8 */
ADD R1, R1, R2    /* R1=R1+R2 */
LD R2, 0(R1)      /* R2=mem[R1] */
```

(*b*) How would a hardware register renaming mechanism remove the name dependencies? Illustrate your answer by providing a version of the code showing the destination and source registers for each instruction after renaming has taken place. Clearly state what free physical registers you assume are available prior to renaming. [4 marks]

(*c*) Why is the removal of name dependencies beneficial within a superscalar processor? [4 marks]

(*d*) In addition to removing name dependencies, for what other purposes may register renaming hardware be used in a superscalar processor? [4 marks]

(*e*) The out-of-order execution of ALU instructions in a superscalar processor is only constrained by the availability of functional units and true data dependencies. Why must the out-of-order execution of memory instructions (e.g. load and store instructions) be constrained further? [4 marks]

8 Denotational Semantics

(*a*) Let D be a poset and let $f : D \to D$ be a monotone function. Give the definition of the *least pre-fixed point*, $\mathit{fix}(f)$, of f. [3 marks]

(*b*) Let D, E be domains and let $p : D \to E$ and $q : E \to D$ be continuous functions.

(*i*) Prove that $\mathit{fix}(q \circ p) \sqsubseteq q\big(\mathit{fix}(p \circ q)\big)$. [4 marks]

(*ii*) Thereby also prove that $p\big(\mathit{fix}(q \circ p)\big) \sqsubseteq \mathit{fix}(p \circ q)$. [4 marks]

Hence,

$$(\star) \qquad \mathit{fix}(q \circ p) = q\big(\mathit{fix}(p \circ q)\big) \quad \text{and} \quad p\big(\mathit{fix}(q \circ p)\big) = \mathit{fix}(p \circ q) \;.$$

(*c*) Let D, E be domains and let $f : (D \times E) \to D$ be a continuous function. Define $f^\dagger : E \to D$ to be the function $f^\dagger \stackrel{\text{def}}{=} \lambda e \in E.\, \mathit{fix}\big(\lambda d \in D.\, f(d, e)\big)$. Show that $f^\dagger$ is continuous. [5 marks]

Analogously, for a continuous function $g : (D \times E) \to E$, let $g^\ddagger : D \to E$ be the function $g^\ddagger \stackrel{\text{def}}{=} \lambda d \in D.\, \mathit{fix}\big(\lambda e \in E.\, g(d, e)\big)$. Then $g^\ddagger$ is continuous.

(*d*) Let D, E be domains and let $f : (D \times E) \to D$ and $g : (D \times E) \to E$ be continuous functions. Define $h : (D \times E) \to (D \times E)$ to be the continuous function $h \stackrel{\text{def}}{=} \lambda (d, e) \in D \times E.\, \big(f(d, e), g(d, e)\big)$.

(*i*) Prove that $\mathit{fix}(h) \sqsubseteq \big(\, \mathit{fix}(f^\dagger \circ g^\ddagger)\, ,\, \mathit{fix}(g^\ddagger \circ f^\dagger)\, \big)$. [2 marks]

[Hint: Recall $(\star)$ above and the pre-fixed point property of $f^\dagger(e)$ and of $g^\ddagger(d)$, for $e = \mathit{fix}(g^\ddagger \circ f^\dagger)$ and $d = \mathit{fix}(f^\dagger \circ g^\ddagger)$.]

(*ii*) Let $\mathit{fix}(h) = (x, y)$.

Prove that $\mathit{fix}(f^\dagger \circ g^\ddagger) \sqsubseteq x$ and $\mathit{fix}(g^\ddagger \circ f^\dagger) \sqsubseteq y$. [2 marks]

[Hint: Recall the pre-fixed point property of $(x, y) = \mathit{fix}(h)$ and the least pre-fixed point property of $f^\dagger(y)$ and of $g^\ddagger(x)$.]

Hence, $\mathit{fix}(h) = \big(\, \mathit{fix}(f^\dagger \circ g^\ddagger)\, ,\, \mathit{fix}(g^\ddagger \circ f^\dagger)\, \big)$.

9 Digital Communication II

(*a*) Reliable transport protocols typically use retransmission timers to decide when a packet that has not been acknowledged should be resent.

Outline the basis on which, and the mechanisms whereby, TCP computes the retransmission timeout value. [10 marks]

(*b*) End-to-end Internet protocols such as TCP compute a viable transmission rate by adjusting a sliding window.

Describe the set of procedures that TCP uses continually to calculate the congestion window. [10 marks]

10 Human–Computer Interaction

You have been assigned responsibility for designing the user interface of a new version of the UK government web facility that is used by taxpayers to make their income tax returns online.

(*a*) You suspect that different taxpayers use the online tax return in a number of quite distinct ways. How might you investigate this, and how could you communicate your findings to the development team? [2 marks]

(*b*) A study of the system data logs shows two surprising patterns of behaviour. Some users repeatedly enter different values and then change them, apparently trying out the results. The other pattern includes long pauses between entering each piece of data, with the same users accessing help and documentation pages in alternation with data entry pages.

(*i*) How might you account for these two patterns? [4 marks]

(*ii*) What usability analysis procedures might you carry out to improve usability of the system for these two classes of user? [4 marks]

(*iii*) What improvements might be experienced by each class as a result? [4 marks]

(*c*) After the project has begun, it is decided that the new design should also support two further classes of user: those who do not want to make a tax return online, but simply need to find out what paper documents to request, and staff within the tax office who need to use the system to review people's tax returns when the taxpayer rings up with a telephone enquiry.

What usability analysis procedures might you use when customising the user interface for these classes of user, and what improvements might they experience as a result? [6 marks]

11 Information Theory and Coding

(*a*) Let X and Y be discrete random variables over state ensembles $\{x\}$ and $\{y\}$ having probability distributions $p(x)$ and $p(y)$, conditional probability distributions $p(x|y)$ and $p(y|x)$, and joint probability distribution $p(x,y)$. Using only these quantities, provide expressions for each of the following:

(*i*) The joint uncertainty $H(X,Y)$ about both random variables. [2 marks]

(*ii*) The uncertainty $H(X|y=b_j)$ about random variable X once it is known that variable Y has taken on a particular value $y=b_j$. [2 marks]

(*iii*) The average uncertainty $H(X|Y)$ remaining about random variable X when Y is known. [2 marks]

(*iv*) The mutual information $I(X;Y)$ between random variables X and Y. (Your answer can use expressions you have defined above.) [2 marks]

(*v*) The union of $H(X|Y)$, $I(X;Y)$, and $H(Y|X)$. [2 marks]

(*b*) A noisy binary communication channel randomly corrupts bits with probability p, so its channel matrix is:

$$\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

(*i*) If the input bit values $\{0,1\}$ are equiprobable, what is the mutual information between the input and output for this noisy channel? [2 marks]

(*ii*) What is the *channel capacity* of this noisy channel? [1 mark]

(*iii*) If an error-correcting code were designed for this noisy channel, what would be the maximum possible entropy of an input source for which reliable transmission could still be achieved? Express your answer in terms of p. [1 mark]

(*iv*) Name the theorem by Shannon that is the basis for the result in (*iii*). [1 mark]

(*c*) Explain why the encoding of continuous signals into sequences of coefficients on Gabor wavelets encompasses, as special cases, both the delta function sampling basis and the Fourier Transform basis. Show how one particular parameter determines where a signal representation lies along this continuum that bridges from delta function sampling to the complex exponential. [5 marks]

12 Optimising Compilers

(*a*) Explain two concepts of a variable being *live* – one related to execution behaviour and one related to the structure of a program. Relate them by implication, and explain their relative ease of computation in a compiler. [4 marks]

(*b*) Explain how live variable analysis can be used to allocate variables to registers by *colouring*. Give and justify an algorithm that performs this colouring, particularly noting how it avoids early decisions causing inconvenient early choices of colour. [5 marks]

Let K_n be the graph of n nodes, each having an edge to each other; let C_n have n nodes, but with n edges arranged to give a cycle; and let S be C_4 with an additional edge forming a diagonal of C_4 seen as a square.

(*c*) What is the minimum number of colours necessary to colour K_n, C_n and S? [3 marks]

(*d*) How many colours does your algorithm require for C_n (if it makes arbitrary choices give both best-case and worse-case)? [2 marks]

(*e*) Give programs that have K_5, C_4 and S as colouring problems for register allocation. [3 marks]

(*f*) Give programs in SSA form (or indicate when this is impossible) for the three graphs in part (*e*). [3 marks]

13 Security

Many critical industries, such as electricity, water, oil and gas, have plant controlled by complex digital systems. These "Supervisory Control and Data Acquisition" (SCADA) systems connect sensors, such as temperature and pressure gauges, with actuators, such as valves and switches, and control rooms. They were originally standalone systems and were thus designed without security mechanisms. However, over the last ten years, they have increasingly acquired Internet connectivity, and now there is serious concern about "cyberterrorism" in the form of online sabotage. As a result, the North American Electric Reliability Corporation (NERC) has ordered critical electricity utilities to protect their networks from 2009 or face substantial fines.

You have been hired by a utility that has several thousand devices on its central sites and several hundred at remote locations (the exact numbers are not known). These devices will disclose data to, or act on data from, anyone who communicates with them using the appropriate protocol.

(*a*) Discuss the advantages and disadvantages of the following protective strategies.

(*i*) Implementing fault-tolerant logic in the control system to identify and isolate faulty sensors. [4 marks]

(*ii*) Using a firewall to isolate the control system network from the corporate network and the Internet. [4 marks]

(*iii*) Authenticating traffic on the control system network by replacing sensors and actuators by, or supplementing them with, devices that can generate and verify message authentication codes. [4 marks]

(*b*) Your customer opts for strategy (*ii*) in respect of central sites and strategy (*iii*) for remote devices. Sketch an overall system design and discuss any residual risk. [8 marks]

14 Specification and Verification I

(*a*) Explain the difference between the *soundness* and *completeness* of a deductive system. [4 marks]

(*b*) Why is soundness generally thought of as more important than completeness? [4 marks]

(*c*) Give an example of a true partial correctness specification that cannot be proved in the version of Hoare logic presented in the lectures. Explain why your example cannot be proved. [4 marks]

(*d*) Devise an instance of the following rule to show that it is not sound.

$$\frac{\{P\}\ C\ \{Q\}}{\{P \wedge R\}\ C\ \{Q \wedge R\}}$$

[4 marks]

(*e*) Devise sufficient conditions on C and R that make the rule above sound. Briefly explain why your conditions ensure soundness. [4 marks]

15 Specification and Verification II

(*a*) List *two* points that compare and contrast the use of Hoare logic and Temporal logic for hardware specification and verification. [4 marks]

(*b*) What is the role of assertions in hardware verification? [4 marks]

(*c*) Explain the difference between *dynamic* and *static* assertion verification. [4 marks]

(*d*) What is the relationship between CTL, LTL and the temporal layer of PSL? What is the modelling layer of PSL? [4 marks]

(*e*) Contrast the importance of logical soundness in bug-finding and in proof of correctness. [4 marks]

END OF PAPER