COMPUTER SCIENCE TRIPOS Part II

Tuesday 5 June 2007 1.30 to 4.30

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
None

1 Comparative Architectures

(*a*) Why is branch prediction so important in modern processors? [2 marks]

(*b*) Why are saturating, two-bit counters used? [2 marks]

(*c*) How does a local branch predictor operate? [2 marks]

(*d*) How does a global branch predictor operate? [2 marks]

(*e*) How may local and global branch predictors be used together? [3 marks]

(*f*) Give an example sequence of branches that will fail to be completely predicted with a branch history of four. [3 marks]

(*g*) How can the compiler assist with branch prediction? [3 marks]

(*h*) Suggest methods for jump prediction. [3 marks]

2 Digital Communication II

Recent advances in consumer electronics have led to the potential demand for *Home Area Networks* (HANs). Given that this would include monitoring and control of the home environment, linking home entertainment (television, radio, MP3 players etc.), and communications (Voice, Data, possibly video) within the home and to and from the outside world, one can consider the relative merits of various forms of HAN.

(*a*) What would be the distinguishing features of an IP-based HAN? [4 marks]

(*b*) What would be the distinguishing features of a cell-switched HAN? [4 marks]

(*c*) What are the comparative merits of the two solutions, considering the application requirements such as guaranteed throughput and jitter; the separation of control and data; the management complexity of the network; and the complexity of the interconnection systems such as switches or routers? [12 marks]

3 Security

A rapidly-growing online crime is *phishing*, in which victims are lured by an e-mail to log on to a website that appears genuine but that actually steals their passwords.

You have been hired by a bank to help them harden their online banking service against phishing attacks.

(*a*) Explain briefly the strengths and weaknesses of the following four possible countermeasures:

(*i*) SSL/TLS client certificates issued to each customer; [4 marks]

(*ii*) a handheld password calculator issued to each customer; [4 marks]

(*iii*) displaying a unique picture to each customer during the login process; [4 marks]

(*iv*) requiring that large payments, or payments to new recipients, be authorised by telephone or SMS as well as online. [4 marks]

(*b*) You are told that the budget will accommodate only *two* of the above options. Which two would you recommend, and why? [4 marks]

4 Advanced Graphics

(*a*) A NURBS curve is defined by control points $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_{n+1}$ and knot vector $(t_1, t_2, \ldots, t_{k+(n+1)})$.

(*i*) State the formulæ for deriving the basis functions $N_{i,k}(t)$. [3 marks]

(*ii*) Graph all seven of the linear basis functions, $N_{i,2}(t)$, for the knot vector $(0, 1, 2, 4, 5, 5, 5, 6, 7)$. [3 marks]

(*iii*) Draw seven points, $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_7$, equi-spaced around a circle and draw the linear NURBS curve defined by those points and the basis functions from part (*ii*). [2 marks]

(*iv*) Derive the formula for and sketch a graph of the basis function $N_{3,3}(t)$ for the knot vector in part (*ii*). [3 marks]

(*b*) Choose *one* of the *Doo–Sabin*, *Catmull–Clark* and *Loop* subdivision schemes. Describe your chosen scheme, including an explanation of how it works in regular regions of the mesh, how it works at and around extraordinary polygons, and how it works at and around extraordinary vertices. [9 marks]

5 Information Retrieval

(*a*) Information retrieval systems are commonly compared against each other in large-scale evaluative conferences. Which difficulties do designers of such evaluations face? Describe *three* difficulties and how they could be addressed. [5 marks]

(*b*) 11-pt avg. precision is an area-based evaluation measure which combines precision and recall.

(*i*) Why are simpler evaluation metrics insufficient for comparative IR evaluation exercises? [3 marks]

(*ii*) Give the formula for 11-pt avg. precision, defining all variables used. [2 marks]

(*iii*) The following table shows the output of an information retrieval system on two queries. You can assume that there are no relevant documents in ranks lower than the top 10 ranks shown. Calculate 11-pt avg. precision for these two queries, using an appropriate interpolation method if necessary, and sketch it in a precision-recall graph. [5 marks]

Rank	Q1	Q2
1	X	–
2	–	X
3	–	–
4	X	X
5	X	–
6	–	X
7	–	X
8	–	–
9	–	X
10	–	X

Output of an IR system on two queries Q1 and Q2; top 10 ranks. Crosses correspond to relevant documents, dashes to irrelevant documents.

(*c*) Large-scale evaluations of information extraction systems use a different setup from IR evaluations, and different evaluation metrics. Describe in detail how you would organise the evaluation of information extraction systems on the task that SNOWBALL solves, i.e. the determination of pairs of company names and headquarters. Which data would you prepare, and which evaluation metrics would you use? Justify your decisions. [5 marks]

6 Specification and Verification I

(*a*) Describe how the meaning of {`T`} `X := Y` {`X=Y`} differs from the meaning of {`Y=y`} `X := Y` {`X=y`}. [2 marks]

(*b*) Is the total correctness specification `[Y=0] X := X/Y [X=X]` true? Justify your answer. [2 marks]

(*c*) Give an expression E such that {`T`} `X :=` E {`X` $= E$} is not true. [2 marks]

(*d*) Explain how specifications containing VDM's hooked variables like $\overleftarrow{X}$ can be translated to specifications that do not use hooked variables. [2 marks]

(*e*) What is the relationship between the provability of verification conditions and the provability of the specification from which they were generated? [2 marks]

(*f*) Define the weakest liberal precondition `wlp(`C`,`Q`)` in higher-order logic. [2 marks]

(*g*) What is the relationship between $\{P\}\ C\ \{Q\}$ and `wlp(`C`,`Q`)`? [2 marks]

(*h*) Explain how $\forall x.\ P(x)$ is represented in terms of λ-abstraction and function application in higher-order logic. [2 marks]

(*i*) Show how to derive a proof rule for the one-arm conditional from the proof rule for the two-arm conditional and the definition of `IF` S `THEN` C as `IF` S `THEN` C `ELSE SKIP`. [4 marks]

7 Specification and Verification II

(*a*) Explain the use of the following when representing circuits in logic:

(*i*) higher-order variables; [2 marks]

(*ii*) conjunction ($\wedge$); [2 marks]

(*iii*) existential quantification ($\exists$). [2 marks]

(*b*) Describe a representation of binary words in logic and define a function that maps a word to the natural number it encodes in binary. [2 marks]

(*c*) Describe how the following components are modelled in higher-order logic:

(*i*) unit-delay; [2 marks]

(*ii*) clocked, edge-triggered D-type register. [2 marks]

(*d*) Let $[t, t']$ denote the closed interval starting at t and ending at t' ($t \leq t'$ and both t and t' are included in the interval). Give definitions in higher-order logic of the predicates

(*i*) Stable

(*ii*) Odd

where: Stable f (t, t') is true if and only if the value of f is constant on the interval $[t, t']$ and Odd f (t, t') is true if and only if f is true an odd number of times in the interval $[t, t']$. [2 + 4 marks]

(*e*) Contrast the simple switch model of transistors with the difference switching model. [2 marks]

8 Information Theory and Coding

(a) Suppose that the following sequence of Yes/No questions was an optimal strategy for playing the "Game of seven questions" to learn which of the letters $\{A, B, C, D, E, F, G\}$ someone had chosen, given that their *a priori* probabilities were known:

"Is it A?"	"No."
"Is it a member of the set $\{B, C\}$?"	"No."
"Is it a member of the set $\{D, E\}$?"	"No."
"Is it F?"	"No."

(i) Write down a probability distribution for the seven letters, $p(A), \ldots, p(G)$, for which this sequence of questions was an optimal strategy. [3 marks]

(ii) With this probability distribution, what was the uncertainty, in bits, associated with each question? [1 mark]

(iii) What is the entropy of this alphabet? [1 mark]

(iv) Now specify a variable length, uniquely decodable, prefix code for this alphabet that would minimise the average code word length. [3 marks]

(v) What is your average coding rate R for letters of this alphabet? [1 mark]

(vi) How do you know that a more efficient code could not be developed? [1 mark]

(b) The signal-to-noise ratio SNR of a continuous communication channel might be different in different parts of its frequency range. Explain how the information capacity C of a noisy continuous communication channel, whose available bandwidth spans from frequency ω_1 to ω_2, may be defined in terms of its signal-to-noise ratio as a function of frequency, $SNR(\omega)$. Define the bit rate for such a channel's information capacity, C, in bits/second, in terms of the $SNR(\omega)$ function of frequency. [5 marks]

(c) An invertible transform generates projection coefficients by integrating the product of a signal with each of a family of functions. In a reverse process, expansion coefficients can be used on those same functions to reproduce the signal. If the functions in question happen to form an orthonormal set, what is the consequence for the projection coefficients and the expansion coefficients? [2 marks]

(d) In the Information Diagram (a plane whose axes are time and frequency), why does the Gabor–Heisenberg–Weyl *Uncertainty Principle* imply that information is *quantised* – i.e. that it exists in only a limited number of independent quanta? [3 marks]

9 Types

(*a*) What is meant by the term *type soundness* in connection with type systems for programming languages? [1 mark]

(*b*) Explain by example how ML-style polymorphism for let-expressions combined with reference types leads to an unsound type system. (As part of your answer you should define the type system, but you need not give a formal definition of how expressions in the language evaluate.) [14 marks]

(*c*) How does Standard ML restrict the typing rule for let-expressions in order to restore type soundness? Why does the restriction render untypeable the example you used in part (*b*)? [5 marks]

10 Computer Systems Modelling

(*a*) Let U be a uniform $(0, 1)$ random variable. Show that for any continuous distribution function $F(x)$, the random variable X defined by

$$X = F^{-1}(U)$$

has the probability distribution function $F(x)$. [4 marks]

(*b*) Use your result in part (*a*) and a uniform $(0, 1)$ random variable, U, to construct random variables for the following two distributions:

(*i*) the uniform (a, b) distribution where a and b are real numbers such that $a < b$; [3 marks]

(*ii*) the exponential distribution $Exp(\lambda)$ with parameter $\lambda > 0$. [3 marks]

(*c*) Suppose that $X_1, X_2, \ldots, X_n$ are independent, identically distributed random variables with mean μ and variance σ^2. Use the central limit theorem to derive an approximate $100(1 - \alpha)$ percent confidence interval for μ. [5 marks]

(*d*) How would you obtain a confidence interval similar to that given in part (*c*) that is exact in the special case where the random variables $X_1, X_2, \ldots, X_n$ have a Normal distribution? [5 marks]

11 VLSI Design

(*a*) Sketch a transistor-level circuit for a 2-input AND gate in static CMOS. [2 marks]

(*b*) Consider the design of a 16-input AND gate in static CMOS.

(*i*) Explain why the 2-input design could not simply be scaled up. [2 marks]

(*ii*) Sketch alternative designs using two and four levels of NAND and NOR gates. [2 × 2 marks]

(*iii*) Use logical effort to estimate the delay of both the designs, assuming that the conducting channel in a pFET has twice the resistance of that in an nFET. [10 marks]

(*iv*) Determine the approximate value of the electrical effort at which their speeds are equal. [2 marks]

12 Human–Computer Interaction

You have been asked to work as a usability consultant, for a company where the development team has created a new version of an existing product. Two important requirements were that the new user interface should be both *efficient* and *intuitive*.

(*a*) How would you interpret *each* of these requirements, in the light of your knowledge of HCI? [1 mark each]

(*b*) The development team claim that their new user interface is already more efficient than the old one. Explain in detail how you could measure whether this claim is justified. [7 marks]

(*c*) The developers also claim that their new version is already more intuitive than the old one. Explain in detail how you could measure whether this claim is justified. [7 marks]

(*d*) What technique could you have used to predict *each* of these measured improvements, if you had been consulted earlier in the design cycle? [2 marks each]

13 Business Studies

(*a*) Give *five* criteria an investor might apply to a start-up proposal. [5 marks]

(*b*) What are the differences between *debt* and *equity* finance? [5 marks]

(*c*) A software start-up company is developing computer games software. They believe their game will have potential market of a million units selling at a retail price of £49.99. They have already raised £1M from Angel investors for 33% of the company, which has been mostly spent on development. They estimate they can complete development and become cash flow positive following initial marketing, but that this will cost a further £1M and take another year. They intend to raise this money by selling further equity.

(*i*) Price this issue. [5 marks]

(*ii*) They receive a letter of intent from a publisher confirming their market estimation and offering 10% royalty on the retail price with £500k recoupable but non-refundable advance (where the publisher will take the first £500k of royalty earned to recoup the advance, but will not demand a refund if the game fails to sell). Should the company take this offer and how does this affect the proposed share offer? [5 marks]

14 E-Commerce

(*a*) Outline *five* business models for e-commerce. [5 marks]

(*b*) Outline *five* methods of valuing an e-commerce business. [5 marks]

(*c*) Give plausible reasons why Google purchased YouTube for $1.65Bn in October 2006, when YouTube had never made a profit. [10 marks]

15 Additional Topics

A large ski resort is organised as a consortium of hundreds of independently-owned lifts, each serving one or more ski slopes. The resort sells a non-transferable, photo-id based "skipass" ticket that lets its owner ride on a specified subset of the lifts, for a specified range of dates, for an unlimited number of rides. Each ticket has a different barcode and is scanned at the entry gate of each lift to decide whether to let the skier through.

The whole resort is about to upgrade from barcode to passive RFID: tickets with embedded RFID tags can be read through the pocket of the skier from a distance of about 5 cm. You have been hired as a security consultant by the resort to oversee the planned changeover.

(a) The operators have requested offline operation: like its barcode-based predecessor, the new system must permit the verification of tickets without requiring individual lifts to have a live connection to the central server all day long (or at all). Justify this request, then design a suitable system architecture and discuss the security implications of this requirement. [6 marks]

(b) Discuss as many ways as possible in which the consortium might be defrauded and suggest appropriate countermeasures. Label each fraud as "made easier by RFID" or "made harder by RFID". Your employers are slightly less interested in frauds that carry over unchanged between barcode and RFID: so, do mention them if you wish but don't over-analyse them. [8 marks]

(c) Small but vocal consumer groups complained about loss of privacy as soon as they heard that the system used RFID. Highlight any privacy threats introduced by RFID in this system. Design an alternative RFID-based architecture offering strong privacy protection for customers. Compare it, from any relevant viewpoints, against the legacy barcode system and against the privacy-indifferent RFID system you designed in part (a). [6 marks]

16 Additional Topics

You are required to design a security system that uses location information about objects and people.

The application scenario is that of a warehouse containing many large boxes. The boxes can be anywhere in the warehouse and are also stacked vertically. Every part of the warehouse is under video surveillance using a very large number of cameras.

When a box goes missing there are two requirements. The first is to produce an audit trail of the movements and location of the box in the warehouse during the previous month. The second is to review quickly the vast amount of recorded video data and select only those scenes that contain the missing box.

(*a*) Describe the architecture of the overall system. [4 marks]

(*b*) Outline the design of an indoor location system that might be used as part of the solution. [6 marks]

(*c*) What are the performance bottlenecks and how does the design deal with scalability issues? [4 marks]

(*d*) What other sensor information could be used to enhance the two application requirements? [6 marks]

END OF PAPER