COMPUTER SCIENCE TRIPOS Part II

Thursday 3 June 2010   1.30 to 4.30

COMPUTER SCIENCE Paper 9

*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 the functions of the space segment, the control segment, and the user segment in the GPS system. [6 marks]

(*b*) An increasing number of mobile telephony handsets are GPS-enabled.

(*i*) Explain why a GPS receiver can take many minutes to locate itself when first turned on. [2 marks]

(*ii*) This length of delay is unacceptable for many mobile handset users. Explain how network operators are able to reduce this delay. [2 marks]

(*iii*) It is not always possible for a GPS receiver to see a sufficient number of GPS satellites to position itself. Describe how a Time Difference of Arrival (TDoA) system would allow a mobile handset to be positioned within a typical cellular network of today. Why is a Time of Arrival (ToA) system inappropriate here? [6 marks]

(*iv*) Indoor location can provide very useful context for sentient computing. Assuming it worked indoors just as it does outdoors, explain why GPS locations might not be good enough for indoor applications. Describe one technique that a mobile handset of the future could potentially use to get reliable indoor location. [4 marks]

2 Advanced Systems Topics

(*a*) The Border Gateway Protocol (BGP) uses two mechanisms to reduce the number of update messages: route-flap damping and *rate-limiting*.

(*i*) Describe the rate-limiting mechanism. [3 marks]

(*ii*) Discuss what might happen if every BGP speaking router in the Internet turned off rate-limiting. [3 marks]

(*b*) Professor Bell claims that BGP could be replaced by a simple *shortest paths* routing protocol. Is Professor Bell correct? Explain. [4 marks]

(*c*) Content Delivery (or Distribution) Networks (CDNs) are becoming an important component of the Internet.

(*i*) What is a CDN? [2 marks]

(*ii*) What impact could the growth of CDNs have on inter-domain routing? [3 marks]

(*d*) The growth of routing tables has led to proposals to make an explicit separation between *identifiers* and *locators*.

(*i*) Explain how this could reduce routing table size. [2 marks]

(*ii*) Explain how *tunnelling* is required by this scheme. [3 marks]

3 Bioinformatics

(*a*) Discuss why the use of spaced seeds in sequence database search is better than the use of consecutive seeds. [5 marks]

(*b*) Discuss the complexity of Sankoff's parsimony method. [5 marks]

(*c*) Describe the four points conditions in phylogeny. [5 marks]

(*d*) Discuss the assumptions of the Gillespie algorithms. [5 marks]

4 Computer Vision

(*a*) Consider an object's surface reflectance map $\phi(i, e, g)$ specifying the amount of incident light reflected towards a camera from each point on the surface, where the angle of the illuminant (a point source) relative to the local surface normal N is i, the angle relative to N of a ray of light re-emitted from the surface is e, and the angle between the emitted ray and the illuminant is g.

(*i*) For what kind of surface is the reflectance map simply $\phi(i, e, g) = \cos(i)$? Name this type of surface and describe its key properties. [3 marks]

(*ii*) For what kind of surface does the reflectance map simplify to $\phi(i, e, g) = 1$ if $i = e$ and both i and e are co-planar with the surface normal N, and $\phi(i, e, g) = 0$ otherwise? Name this type of surface and describe its key properties. [3 marks]

(*iii*) For what kind of surface does the reflectance map depend only on the ratio of the cosines of the angles of incidence and emission, $\cos(i)/\cos(e)$, but not upon their relative angle g nor upon the surface normal N? Give an example of such an object, and explain the consequence of this special reflectance map for the object's appearance. [3 marks]

(*iv*) For the more general class of object surfaces described by the following reflectance map, with $0 \leq s \leq 1$,

$$\phi(i, e, g) = \frac{s(n+1)(2\cos(i)\cos(e) - \cos(g))^n}{2} + (1-s)\cos(i)$$

which term is the specular component, and which term is the matte component? What fraction of light is emitted specularly, and what does the parameter n represent? Describe the behaviour of parameters s and n around regions of a person's face where the skin might be oily or sweaty and places where it is not, and how this impacts recognition. [4 marks]

(*b*) Using the second finite difference operator $\boxed{-1 \mid 2 \mid -1}$ for edge detection in an image, show how the pixel values in the top row are changed by discrete convolution with this operator: insert the output values in the bottom row.

...	0	0	0	0	5	5	5	5	5	0	0	0	0	...

[3 marks]

(*c*) Show how Bayesian inference enhances face recognition when a face contains highly distinctive features, as exploited by caricature. Suppose some facial feature x is unusual so its probability $P(x)$ is small, and that for each k^{th} face described as class C_k we know the class-conditional likelihood $P(x|C_k)$ of observing this unusual feature, but *a priori* all the classes are equiprobable. Use Bayes' Rule to show how the correct classification of face C_k given its unusual feature x acquires higher probability $P(C_k|x)$. [4 marks]

5 Denotational Semantics

Indicate whether the following statements are true or false. Provide an argument for each answer. You may use standard results provided that you state them clearly.

(*a*) For all PCF-types τ and closed PCF-terms M_1 and M_2 of type τ,

if $[\![M_1]\!] = [\![M_2]\!]$ in $[\![\tau]\!]$ then $M_1 \cong_{\mathrm{ctx}} M_2 : \tau$ [6 marks]

(*b*) For all PCF-types τ and closed PCF-terms M_1 and M_2 of type τ,

if $M_1 \cong_{\mathrm{ctx}} M_2 : \tau$ then $[\![M_1]\!] = [\![M_2]\!]$ in $[\![\tau]\!]$ [7 marks]

(*c*) For all closed PCF-terms M_1 and M_2 of type $nat \to nat$,

if $M_1 \cong_{\mathrm{ctx}} M_2 : nat \to nat$ then $[\![M_1]\!] = [\![M_2]\!]$ in $(\mathbb{N}_\perp \to \mathbb{N}_\perp)$ [7 marks]

6 Digital Communication II

(*a*) A network consists of a central hub router, with links (spokes) to a set of routers at outlying sites. A network designer is trying to choose which routing protocol to deploy.

(*i*) Illustrate the operation of Link State and Distance Vector routing algorithms over this simple topology. [6 marks]

(*ii*) What are the overall packet counts for each protocol in the whole network from turning the system on until all routers know where all destinations are? [2 marks]

(*b*) Another network consists of a set of sites with routers, connected in a long line.

(*i*) Illustrate the operation of Distance Vector routing when one link at one end fails. [6 marks]

(*ii*) What are **two** common techniques to improve the performance of the protocol in these scenarios? [4 marks]

(*c*) Why do path vector protocols not have this problem? [2 marks]

7 Digital Signal Processing

(*a*) Make the following statements correct by changing one word or number. (Negating the sentence is not sufficient.)

(*i*) An absolutely summable discrete sequence will have in the corresponding z-transform plane at $z = 1$ a positive value. [1 mark]

(*ii*) A memory-less system depends only on the next input value. [1 mark]

(*b*) Define the convolution operator on discrete sequences. [2 marks]

(*c*) Prove that convolution of discrete sequences is an associative operation. [6 marks]

(*d*) Given samples $x_n = x(t_\mathrm{s} \cdot n)$ for all integers n, where $x(t)$ is a continuous signal whose Fourier transform has non-zero values only at frequencies f with $f_\mathrm{l} < |f| < f_\mathrm{h}$,

(*i*) under which condition can the original waveform $x(t)$ be reconstructed; [4 marks]

(*ii*) and how can this be done? [6 marks]

8 Information Theory and Coding

(*a*) Show that for an alphabet of unlimited size, it is possible to create a uniquely decodable prefix code for its letters such that their average codeword length is no more than 2 bits, provided that the alphabet letters obey a certain probability distribution. (No letter has probability 0; the letters' probabilities all sum up to 1; and "average codeword length" weights each codeword length by the probability of occurrence of the letter.) Give the necessary descending sequence of letter probabilities, and a uniquely decodable codeword with the prefix property for each letter, and show that the average codeword length remains ≤ 2 bits regardless of how large the alphabet may be. [6 marks]

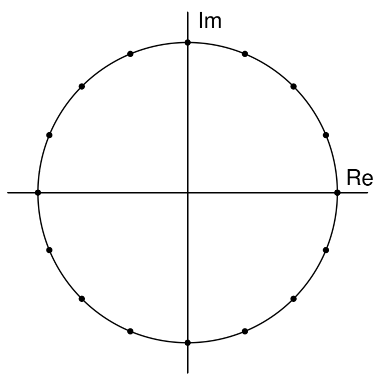
Hint: you may invoke without proof the following series limit:

$$\lim_{N\to\infty} \sum_{n=1}^{N} \frac{1}{2^n} \log_2(2^n) = 2$$

(*b*) Why is the JPEG2000 image compression protocol so much better than the original JPEG in delivering good results even at severe compression rates? What are the main differences in the basis functions or coding wavelets used, what reconstruction artefacts are thereby avoided, and why? [4 marks]

(*c*) The signal-to-noise ratio SNR of a continuous communication channel can be different in different parts of its frequency range. For example, the noise might be predominantly high frequency hiss, or low frequency rumble. Explain how the information capacity C of a noisy continuous communication channel, whose available band is the frequency range from ω_1 to ω_2, may be defined in terms of its signal-to-noise ratio as a function of frequency, SNR(ω). Define the bit rate for this noisy channel's information capacity, C, in bits/second, in terms of this SNR(ω) function of frequency. [6 marks]

(*d*) Suppose a discrete data sequence $\{g_n\}$ consisting of 16 data points is Fourier transformed using an FFT algorithm. The fourth frequency component is a complex exponential that evolves four cycles over the course of the 16-point data sequence. Using a unit circle diagram in the complex plane as shown here, explain why the 16 complex multiplications implicitly needed to compute any Fourier coefficient of this data can be reduced to just 4 multiplications for this coefficient, by first adding together certain data points in $\{g_n\}$.



[4 marks]

9 Natural Language Processing

The following is a fragment of an annotated corpus that could be used for training and testing a pronoun resolution algorithm.

```
<s>Adapting <e id="1" pron="f">this first verse</e>,
<e id="2" pron=f>Burns</e> did <e id="3" pron="f">something
unforeseen</e>.<s><e id="4" pron="t">He</e> obscured and
mystified <e id="5" pron="f"> <e id="6" pron="t">its</e>
sentiments</e>

Key: 4=2, 6=1
```

`<e></e>` delimits a referring expression, `pron` indicates whether the expression is a pronoun to be resolved, and `<s>` marks sentence boundaries. The key gives the antecedents of the pronouns.

(*a*) Using this fragment as an example, illustrate how pronoun resolution can be treated as supervised classification using a Naïve Bayes classifier. You should include details of possible features, explaining why they are relevant and outlining how they might be derived where appropriate. Details of the Naïve Bayes classifier are not required. [15 marks]

(*b*) Why might better performance be achieved in pronoun resolution if a discourse model that records how other pronouns have been resolved is utilised? [3 marks]

(*c*) What is meant by the term *baseline* in the context of NLP evaluation? Suggest a possible baseline algorithm for the pronoun resolution task. [2 marks]

10 Optimising Compilers

A CPU has the following two features:

1. Every arithmetic instruction (as well as the special “compare” instruction) sets condition codes based on the value of the result, and subsequent conditional instructions can test these.

2. If you execute an instruction that attempts to load from or store to a memory address that is not a multiple of the word-size, the CPU generates an interrupt.

You are concerned with decompilation from the target instructions of this machine back into a high-level programming language.

(*a*) (*i*) What issues are raised by the condition codes, and what effect would they have on a naïve decompilation? [5 marks]

(*ii*) What optimisation techniques could you use to improve matters, and what (if any) are the limitations of your approach? [5 marks]

(*b*) Direct decompilation would need to prefix every memory access with a test to see whether the address being used was properly aligned. This would lead to bulky code that was hard to read and had poor performance. It is expected that, in the program that is being decompiled, unaligned memory accesses are very rare. You are expected to decompile into portable code in the high-level language that may not rely on any treatment of unaligned addresses on the fresh computer you compile for. Again, what optimisation techniques could you apply? [10 marks]

11 Quantum Computing

(*a*) Which of the following are possible states of a qubit?

(*i*) $\frac{1}{\sqrt{2}}(|1\rangle + |0\rangle)$

(*ii*) $0.6|0\rangle + 0.4|1\rangle$

(*iii*) $0.8|0\rangle - 0.6e^{3i\pi/4}|1\rangle$

(*iv*) $\frac{\sqrt{3}}{2}i|0\rangle - \frac{1}{2}|1\rangle$

[1 mark each]

For each valid state among the above, give the probabilities of observing $|0\rangle$ and $|1\rangle$ when the system is measured in the standard computational basis.

[4 marks]

(*b*) Suppose a two-qubit system is in the state $0.8|00\rangle + 0.6|11\rangle$. A Pauli X gate (i.e. a NOT gate) is applied to the second qubit, and a Hadamard gate is applied to the first qubit.

(*i*) What is the new state of the system? [2 marks]

(*ii*) What are the probabilities of the possible outcomes if both qubits are now measured? [2 marks]

(*c*) Suppose we have an algorithm which, given a blackbox computing a periodic function f with range $\{0, \ldots, N-1\}$, determines the period of f. Moreover, the algorithm runs in time $(\log N)^2$. Explain how this would enable us to have an efficient (i.e. polynomial-time) algorithm for factoring numbers. [8 marks]

12 System-on-Chip Design

(*a*) What is the critical path of a synchronous clock domain and how does it affect performance? [2 marks]

(*b*) Name **two** tools that can be used for locating the critical paths of a system on chip before fabrication. Which is faster or better and why? [4 marks]

(*c*) Explain how pipelining can alter performance. What are its disadvantages? [4 marks]

(*d*) How does pipelining alter the state encoding for a sub-system? [2 marks]

(*e*) Name a tool that can be used to formally check whether a minor design change has altered the behaviour of a subsystem where the state encoding has not been changed. Name another tool for use when the state encoding may have been changed. [2 marks]

(*f*) Using any combination of diagrams, SystemC or RTL, sketch **two** implementations for a design of your choosing that use a different number of flip-flops to achieve the same functionality. Explain whether the reports from each of your two tools from part (*e*) might differ for your two implementations. [6 marks]

13 Specification and Verification II

REG is a unit-delay register initialising to some unknown value, the combinational devices TEST and STEP compute the functions test (which returns a Boolean) and step, respectively, and MUX is a multiplexer. These are specified by:

$$
\begin{array}{lcl}
\mathsf{REG}(\mathit{in},\ \mathit{out}) & = & \forall t.\ \mathit{out}(t{+}1) \ = \ \mathit{in}\ t \\
\mathsf{TEST}(\mathit{in},\ \mathit{out}) & = & \forall t.\ \mathit{out}\ t \ = \ \mathsf{test}(\mathit{in}\ t) \\
\mathsf{STEP}(\mathit{in},\ \mathit{out}) & = & \forall t.\ \mathit{out}\ t \ = \ \mathsf{step}(\mathit{in}\ t) \\
\mathsf{MUX}(\mathit{select},\ \mathit{in}_1,\ \mathit{in}_2,\ \mathit{out}) & = & \forall t.\ \mathit{out}\ t \ = \ \texttt{if}\ \mathit{select}\ t\ \texttt{then}\ \mathit{in}_1\ t\ \texttt{else}\ \mathit{in}_2\ t
\end{array}
$$

These components can be used to design a device, INIT, that attempts to initialise to an output value satisfying test by applying step zero or more times to the initial value of REG (i.e. the value output at time 0) until test yields $\texttt{T}$. INIT indicates that initialisation has succeeded by outputting $\texttt{T}$ on the output line *done*; it simultaneously outputs the value found by repeatedly applying step on the output line *out*. Formally, INIT is specified by two properties:

$$
\begin{array}{lcl}
\mathsf{INIT}(\mathit{done}, \mathit{out}) \ \Rightarrow \ \forall t.\ \mathit{done}\ t & = & \mathsf{test}(\mathit{out}\ t) \\
\mathsf{INIT}(\mathit{done}, \mathit{out}) \ \Rightarrow \ \forall t.\ \mathit{out}(t{+}1) & = & \texttt{if}\ \mathit{done}\ t\ \texttt{then}\ \mathit{out}\ t\ \texttt{else}\ \mathit{step}(\mathit{out}\ t)
\end{array}
$$

(*a*) Give a design for INIT in the form of a circuit diagram using REG, TEST, STEP and MUX. [6 marks]

(*b*) Carefully explain why your design meets its specification. [4 marks]

(*c*) Write down a formal model of your design. [4 marks]

(*d*) Outline how you could prove that your design meets its specification (you need not give a detailed proof, but you should provide evidence that you know how to produce such a proof, and what the main steps would be). [6 marks]

14 Topics in Concurrency

(*a*) Carefully state and prove Tarski's fixed-point theorem for the maximum fixed point $\nu X.\varphi(X)$ of a monotonic function φ on the powerset $\mathcal{P}(A)$ of a set A. [6 marks]

(*b*) Let $S \subseteq A$. Prove

$$S \subseteq \nu X.\varphi(X) \;\Leftrightarrow\; S \subseteq \varphi(\nu X.(S \cup \varphi(X)))$$

[7 marks]

(*c*) Describe the syntax and semantics of the modal mu-calculus. Describe a model checking algorithm which uses the fact in part (*b*) to decide whether an assertion of the modal mu-calculus holds of a state in a finite transition system. [7 marks]

15 Types

Consider the following type and expressions of the Polymorphic Lambda Calculus (PLC):

$$\begin{aligned} nat &= \forall\alpha(\alpha \to (\alpha \to \alpha) \to \alpha) \\ Z &= \Lambda\alpha(\lambda x : \alpha(\lambda f : \alpha \to \alpha\,(x))) \\ S &= \lambda y : nat(\Lambda\alpha(\lambda x : \alpha(\lambda f : \alpha \to \alpha\,(f(y\,\alpha\,x\,f))))) \end{aligned}$$

(a) What are the types of Z and S? [2 marks]

(b) Show that there is a closed PLC expression I of type

$$\forall\alpha(\alpha \to (\alpha \to \alpha) \to nat \to \alpha)$$

satisfying the following beta-conversions:

$$\begin{aligned} I\,\alpha\,x\,f\,Z &=_\beta x \\ I\,\alpha\,x\,f\,(S\,y) &=_\beta f(I\,\alpha\,x\,f\,y) \end{aligned}$$

[4 marks]

(c) For each natural number $n \in \mathbb{N} = \{0, 1, 2, \ldots\}$, let $S^n Z$ be the PLC expression given by

$$\begin{aligned} S^0 Z &= Z \\ S^{n+1} Z &= S(S^n Z) \end{aligned}$$

What is the beta-normal form of $S^0 Z$, of $S^1 Z$, of $S^2 Z$, and in general, of $S^n Z$? [4 marks]

(d) (i) Using part (b), or otherwise, show that there is a closed PLC expression A of type $nat \to nat \to nat$ that represents addition of natural numbers, in the sense that $A(S^m Z)(S^n Z) =_\beta S^{m+n} Z$ holds for all $m, n \in \mathbb{N}$. [Hint: recall the primitive recursive definition of addition.] [5 marks]

(ii) Show that $M = \lambda y : nat(I\,nat\,Z\,(A\,y))$ (with A is as in part (i)) represents multiplication of natural numbers, in the sense that $M(S^m Z)(S^n Z) =_\beta S^{mn} Z$ holds for all $m, n \in \mathbb{N}$. [5 marks]

END OF PAPER