UNIVERSITY OF CAMBRIDGE

COMPUTER SCIENCE TRIPOS Part II

Thursday 4 June 2015 1.30 to 4.30 pm

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
Rough work pad

SPECIAL REQUIREMENTS
Approved calculator permitted

1 Bioinformatics

(*a*) Define the Longest Common Subsequence (LCS) problem between two strings and find a solution for the case of the two strings:

ACGT and GGTTTAAGCCGT

[4 marks]

(*b*) Discuss the limitations of the Nussinov algorithm for RNA folding prediction. [4 marks]

(*c*) Which algorithm would you choose for clustering sequence and gene expression datasets and why? [3 marks]

(*d*) Discuss the complexity of the Markov Clustering (MCL) algorithm. [4 marks]

(*e*) Discuss the assumptions for the use of the Gillespie algorithm in the study of reaction networks and comment on the algorithm's complexity in terms of the number of reactions. [5 marks]

2 Computer Systems Modelling

Consider a birth death model with birth rates λ_n in states $n = 0, 1, \ldots$ and death rates μ_n in states $n = 1, 2, \ldots$.

(*a*) State the detailed balance conditions for the equilibrium probability p_n of being in state n, for $n = 0, 1, \ldots$ and explain how they are derived. [2 marks]

(*b*) Using the detailed balance conditions derive an expression p_n and state any condition needed to ensure that the equilibrium distribution exists. [4 marks]

(*c*) Now consider the M/M/1 queue with arrival rate λ and service rate μ and explain how it can be modelled as a birth death model. [2 marks]

(*d*) For the M/M/1 queue derive the form of p_n, the equilibrium distribution of the number of customers present, and state any conditions needed to ensure the existence of the equilibrium distribution. Derive the mean value of the equilibrium distribution. [4 marks]

(*e*) Now consider a M/M/1 model with the modification that customers waiting in the queue are impatient and will only wait for an exponentially distributed amount of time with rate parameter θ before departing the queue without service.

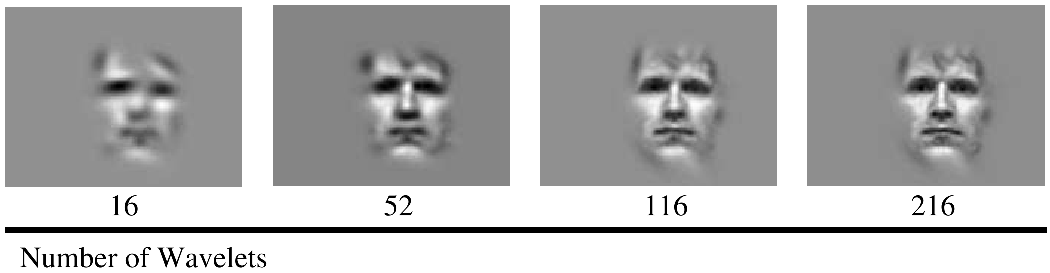
(*i*) Explain how you could modify your birth death model in this situation and write an expression for p_n. [4 marks]

(*ii*) Let α be the probability that a customer receives service and derive an expression for α using Little's law applied to the server. You may leave your expression in terms of the value p_0. [4 marks]

3 Computer Vision

(*a*) In human vision, photoreceptors (cones) responsible for colour are numerous only near the fovea, mainly in the central ± 10 degrees. High spatial resolution likewise exists only there. So then why does the visual world appear to contain colour information everywhere in the field of view? Why does it also seem to have uniform spatial resolution? Why does the world appear stable despite all our eye movements? Discuss the implications for computer vision principles that might be drawn from these observations. [5 marks]

(*b*) Explain why such a tiny number of 2D Gabor wavelets as shown in this sequence are so efficient at representing faces, and why such wavelet-based encodings are able to deliver impressive accuracy performance in "appearance-based" algorithms for face recognition.



[5 marks]

(*c*) Explain the "receptive field" concept in vision, and what is accomplished by the lateral signal flows within both of the plexiform layers of the mammalian retina, in terms of spatial and temporal image processing and coding.

[5 marks]

(*d*) Machine learning plays an increasingly important rôle in computer vision in a strand of work that may be called "learning to see". Compare and contrast discriminative methods with generative methods for constructing classifiers in computer vision. [5 marks]

4 Denotational Semantics

(*a*) (*i*) Define the contextual-equivalence relation $\Gamma \vdash M \cong_{\text{ctx}} M' : \tau$ for pairs of PCF terms M, M', PCF types τ, and PCF type environments Γ. [3 marks]

(*ii*) For PCF terms M and N with respective typings $\Gamma \vdash M : \tau \to \alpha$ and $\Gamma \vdash N : \alpha \to \sigma$, let $N \circ M$ be the PCF term $\mathbf{fn}\ x : \tau.\ N\,(M\,x)$, where $x \notin \mathrm{dom}(\Gamma)$, with typing $\Gamma \vdash N \circ M : \tau \to \sigma$.

State whether or not if $\Gamma \vdash M \cong_{\text{ctx}} M' : \tau \to \alpha$ and $\Gamma \vdash N \cong_{\text{ctx}} N' : \alpha \to \sigma$ then $\Gamma \vdash N \circ M \cong_{\text{ctx}} N' \circ M' : \tau \to \sigma$. Justify your answer. [5 marks]

(*b*) By considering the countable chain of functions $(P_n)_{n\in\mathbb{N}}$ in the function domain $(\mathbb{N}_\perp \to \mathbb{B}_\perp)$ given by

$$P_n(k) \stackrel{\text{def}}{=} \begin{cases} \mathit{false} & \text{if } k \in \mathbb{N} \text{ and } k < n \\ \perp & \text{otherwise} \end{cases} \qquad (k \in \mathbb{N}_\perp)$$

or otherwise, show that the function ε from $(\mathbb{N}_\perp \to \mathbb{B}_\perp)$ to $\mathbb{B}_\perp$ given by

$$\varepsilon(P) \stackrel{\text{def}}{=} \begin{cases} \mathit{true} & \text{if } \exists\, n \in \mathbb{N}.\ P(n) = \mathit{true} \\ \mathit{false} & \text{if } \forall\, n \in \mathbb{N}.\ P(n) = \mathit{false} \\ \perp & \text{otherwise} \end{cases} \qquad \big(P \in (\mathbb{N}_\perp \to \mathbb{B}_\perp)\big)$$

is not continuous. Argue as to whether or not ε is definable by a closed term of type $(\mathit{nat} \to \mathit{bool}) \to \mathit{bool}$ in both PCF and PCF+por. [5 marks]

(*c*) Let M be the PCF+por term

$$\begin{array}{l} \mathbf{fn}\ f : (\mathit{nat} \to \mathit{bool}) \to \mathit{bool}\,. \\ \quad \mathbf{fn}\ P : \mathit{nat} \to \mathit{bool}\,. \\ \qquad \mathbf{por}\Big(P\,\mathbf{0}\,,\ f\,\big(\mathbf{fn}\ n : \mathit{nat}.\ P\,(\mathbf{succ}(n))\big)\Big) \end{array}$$

Give an explicit description of $[\![\, \mathbf{fix}(M)\,]\!] \in \big((\mathbb{N}_\perp \to \mathbb{B}_\perp) \to \mathbb{B}_\perp\big)$. [7 marks]

5 Digital Signal Processing

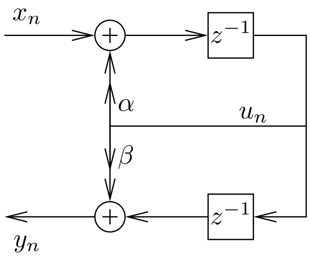
(*a*) You have been asked to design a long-wave radio receiver that can simultaneously monitor two radio signals at 60 ± 10 kHz and 100 ± 10 kHz, that is two stations with 20 kHz bandwidth each. Analog filters in your antenna amplifier suppress all signals outside those two bands.

What is the lowest possible sampling frequency that you can use for a *single* time-domain discrete sequence that records signals from these two bands simultaneously and unambiguously if you use

(*i*) base-band sampling; [2 marks]

(*ii*) IQ sampling. [2 marks]

(*b*) Consider the following digital filter with two multipliers:



(*i*) State the equations that define the elements of the output sequence $\{y_n\}$ and the intermediate sequence $\{u_n\}$ in terms of other values from $\{x_n\}$, $\{y_n\}$ or $\{u_n\}$. [4 marks]

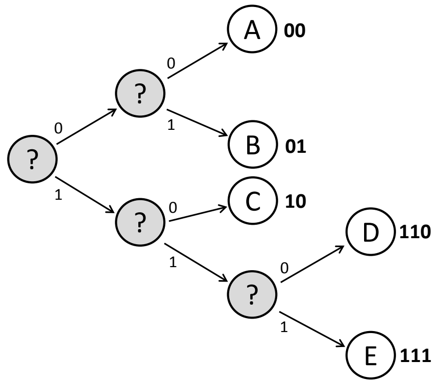
(*ii*) Convert these equations into equivalent equations for the z-transforms $X(z)$, $Y(z)$ and $U(z)$ of these three discrete sequences, and then solve for $U(z)$ and $Y(z)$. [4 marks]

(*iii*) What is the z-transform $H(z)$ of the impulse response of this digital filter? [4 marks]

(*iv*) Draw the block diagram of an equivalent Direct Form I filter. [4 marks]

6 Information Theory and Coding

(*a*) Huffman trees enable construction of uniquely decodable prefix codes with optimal codeword lengths. The five codewords shown here for the alphabet {A,B,C,D,E} form an instantaneous prefix code.



(*i*) Give a probability distribution for the five letters that would result in such a tree.

(*ii*) Calculate the entropy of that distribution.

(*iii*) Compute the average codeword length for encoding this alphabet, and relate your results to the Source Coding Theorem.

[3 × 2 marks]

(*b*) What does it mean for a function to be "self-Fourier"? Show that the Gaussian function is self-Fourier. Name two other functions of importance in information theory that are self-Fourier, and in both cases mention a topic or theorem exploiting this property. [6 marks]

(*c*) (*i*) In the FFT algorithm, if a discrete data sequence consists of N sample values (nominally N is some power of 2), what complex number is the *primitive* N^{th} *root of unity* which, raised to various powers, generates all the complex numbers needed to perform a discrete Fourier transform?

[2 marks]

(*ii*) If all the N^{th} roots of unity are known, by what mechanism are sequences of them selected that are needed for the k^{th} frequency component?

[2 marks]

(*d*) Define the Kolmogorov algorithmic complexity K of a string of data. What approximate relationship is expected between K and the Shannon entropy H for the same source? Give a reasonable estimate of the Kolmogorov complexity K of a fractal, and explain why it is reasonable. [4 marks]

7 Mobile and Sensor Systems

NewZee Natural Reserve is experimenting with mobile and sensing technology to track the behaviour of its animals and the movement of its customers. The Reserve is located in an area of generally good cellular coverage. The Kiwi Paradise is a special enclosure 10km square, within the reserve, where kiwi birds are bred and kept. The enclosure is in the most remote part of the reserve and cellular connectivity is not present in the area. Kiwi birds are non-flying birds, slightly bigger than chickens.

(*a*) Describe the hardware details of a system of your choice to monitor the interactions and mobility of the kiwi birds in the Kiwi Paradise enclosure. State all your assumptions. [4 marks]

(*b*) Describe the software (MAC and Routing) aspects of the system defined in answer (*a*) as well as the energy consumption implications. [7 marks]

(*c*) Kiwi birds are threatened by non-flying predators (such as possums). Describe an extension to your system which is able to detect the intrusion of the predators in the enclosure and raise alerts. Justify your choice. [5 marks]

(*d*) The reserve would like to monitor the movement and behaviour of the customers on its grounds. As indicated above, apart from the Kiwi Paradise enclosure, the reserve is under good cellular coverage. Describe how the customers' smart phones can be used to understand customers' behaviour in the reserve (excluding the Kiwi Paradise enclosure). [4 marks]

8 Natural Language Processing

(*a*) The following text is from 'The Tale of Peter Rabbit' by Beatrix Potter (slightly modified):

> Peter never stopped running or looked behind him till he got home to the big fir-tree.
>
> He was so tired that he flopped down upon the nice soft sand on the floor of the rabbit-hole and shut his eyes. His mother was busy cooking; she wondered what he had done with his clothes. It was the second little jacket and pair of shoes that Peter had lost in a week!

Describe **six** features that are used in pronoun resolution algorithms using classifiers. For each feature, explain the range of values it can take, giving illustrative examples from the text above. [8 marks]

(*b*) Using the text from (*a*) to provide examples, explain what information you would ask a human annotator to give for pronoun resolution, and show how this relates to the training and test data needed for an algorithm using a classifier. [4 marks]

(*c*) The general problem of anaphora resolution includes finding antecedents for definite expressions, such as *the jacket* having *his clothes* as an antecedent. Outline how one might exploit semantic relations from WordNet as part of an extension of the classifier approach to deal with definite expressions. What limitations might WordNet have for this task? [8 marks]

9 Optimising Compilers

(*a*) Explain how a classical dataflow analysis, such as live variable analysis, can be alternatively seen as first walking over the flowgraph and emitting a set of equality constraints, and later solving these constraints to find their least solution. Your answer should make clear the syntactic form of the constraints, the space of values on which equality is to be performed, and what it means to be 'least'. You do not need to explain how the resulting constraints are solved. [4 marks]

(*b*) Reformulate your answer to part (*a*) so that the the flowgraph walker now produces only inequality constraints—but these replacement constraints should have the same least solution as your equality constraints given in part (*a*). [2 marks]

(*c*) Your answer to part (*b*) may have involved a *big union* or *big intersection* syntactic form, e.g. $\bigcup_{s \in S}(\ldots s \ldots)$. Either explain why it did not need to do so, or show how the flowgraph walker may alternatively emit inequality constraints which do not require use of *big union* or *big intersection* but whose solution is identical to that in parts (*b*) and (*a*). Again make clear the syntactic form of your constraints. Illustrate your answer by giving the (inequality-only) constraints emitted for liveness analysis of the flowgraph for the naively translated C code:

```
entry: x = rand();
       y = rand();
       z = 2 + y;
       if (x>10)
           return z;
       /* unspecified code having liveness L6 */
```

[5 marks]

(*d*) Summarise the purpose of *either* 0-CFA *or* Andersen's analysis. Also explain, for your chosen analysis, to what extent it fits the idea of "walk a data structure emitting constraints" and what form of information the analysis computes. You do *not* need to explain in detail the form of generated constraints. [4 marks]

(*e*) Both the analyses named in part (*d*) need to generate constraints more general than those in part (*b*), in particular constraints of the form $t \in S \Longrightarrow T \supseteq U$. Explain how this arises for your chosen analysis. [5 marks]

10 System-on-Chip Design

(*a*) The key for a modern motor car contains a computer, a battery and a radio transmitter. The key has three push buttons. It sends a binary sequence for each key press.

(*i*) Sketch a basic block diagram of a feasible implementation. Show all the major components and indicate how many pieces of silicon might be used. Briefly discuss whether a standard part or a custom SoC should be used in this application. [4 marks]

(*ii*) Explain the likely internal processing steps for the key when a button is pressed with emphasis on how the computer does its I/O. [4 marks]

(*iii*) How would you predict the likely battery life of the key before prototype manufacture? [4 marks]

(*b*) (*i*) Give a pair of synchronous circuits for an RTL module that have identical observable external behaviour but that differ in energy use. Either use RTL or a draw a diagram. Each should have at least one flip-flop and just a few combinational gates and no redundant logic. [3 marks]

(*ii*) State with justification what form of equivalence checker would be needed to determine that your circuits are equivalent. [2 marks]

(*iii*) Say why your pair of circuits use different amounts of energy. Give answers for both static and dynamic power. [3 marks]

11 Topical Issues

A single-floor shopping centre wishes to locate the mobile phones of centre users without their explicit consent and without any modification to their handsets. To do so they deploy receivers throughout the centre to capture the registration messages each handset sends to remain registered on a mobile network.

For this question you should assume these registration messages are sent regularly and contain the handset identifier as a payload.

(*a*) Discuss the applicability of the following positioning techniques to this system. Your answers should describe the principles of each technique and list any requirements on the receivers for each technique deemed applicable.

(*i*) Time of Flight/Arrival (ToF/ToA); [3 marks]

(*ii*) Time Difference of Arrival (TDoA); [4 marks]

(*iii*) Angle of Arrival (AoA); [3 marks]

(*iv*) Received Signal Strength (RSS). [2 marks]

(*b*) Describe a modification to the registration message that could be used by the mobile network to avoid a handset being tracked around the centre. [3 marks]

(*c*) Some centre users are willing to consent to the tracking. Propose a system that can augment the location system to provide more accurate and robust tracking when using handset sensors and radios that are typically found on smartphones. Discuss the strengths and weaknesses of your approach with particular reference to practical issues. [5 marks]

12 Topics in Concurrency

(*a*) Define when a relation R is a *(strong) bisimulation* and define *bisimilarity*. [3 marks]

(*b*) Let the pure CCS processes P and Q be

$$P \stackrel{\text{def}}{=} a.Q \qquad\qquad Q \stackrel{\text{def}}{=} b.P$$

(*i*) Show that: $a.b.P + b.a.Q \sim P + Q$ [4 marks]

(*ii*) Use the local model checking algorithm to show that

$$P \vdash \nu X(\langle\cdot\rangle T \wedge [\cdot]X)$$

reduces to **true** [4 marks]

(*iii*) Given that $a.(b.nil + Q) \vdash \nu X(\langle\cdot\rangle T \wedge [\cdot]X)$ reduces to **false**, are P and $a.(b.nil + Q)$ bisimilar? State carefully but do not prove any results upon which your answer relies. [3 marks]

(*c*) Explain how bisimilarity $\sim$ is a greatest fixed point. State carefully but do not prove any results upon which your answer relies. [6 marks]

13 Types

(*a*) Give the rules for typing the ML expressions for function abstraction and application, reference creation, dereferencing and assignment. [5 marks]

(*b*) What is the *value-restricted* form of the typing rule for let-expressions in ML? Briefly explain what is the purpose of imposing the value restriction on the typing rule. [5 marks]

(*c*) Which of the following typing judgements are provable in the ML type system with the value-restricted typing rule for let-expressions? Justify your answer in each case.

(*i*) $\{\} \vdash \mathtt{let}\, r = \mathtt{ref}\,(\lambda x(x\,\mathtt{true}))\,\mathtt{in}\,(r := \lambda z(\mathtt{true})) : \mathit{unit}$

(*ii*) $\{\} \vdash \mathtt{let}\, r = \mathtt{ref}\,(\lambda x(x\,\mathtt{true}))\,\mathtt{in}\,(!r\,(\lambda y(r := \lambda z(\mathtt{true})))) : \mathit{unit}$

(*iii*) $\{\} \vdash \mathtt{let}\, r = (\lambda x(\lambda y(x)))\,\mathtt{true}\,\mathtt{in}\, r\,(r\,()) : \mathit{bool}$

(*iv*) $\{\} \vdash \mathtt{let}\, r = \lambda y(\mathtt{true})\,\mathtt{in}\, r\,(r\,()) : \mathit{bool}$

[10 marks]

END OF PAPER