CST2
COMPUTER SCIENCE TRIPOS Part II

Monday 5 June 2023 14:00 to 17:00

COMPUTER SCIENCE Paper 8

*Answer **five** questions.*

*Submit each question answer in a **separate** PDF. As the file name, use your candidate number, paper and question number (e.g., `1234A-p8-q6.pdf`). Also write your candidate number, paper and question number at the start of each PDF.*

You must follow the official form and conduct instructions for this online examination

1 Advanced Computer Architecture

(*a*) The Branch Target Buffer (BTB) stores information about previously encountered branch instructions. A simple design would store the entire address of the branch instruction and an entire target address. How could the number of bits stored in each entry of the BTB be significantly reduced and what trade-offs are involved? [4 marks]

(*b*) The BTB and branch predictor are normally accessed on every clock cycle regardless of whether the instruction being fetched is a branch. Assuming a scalar processor, outline one idea that could be used to help reduce these unnecessary accesses? [4 marks]

(*c*) Modern superscalar processors are able to support hundreds of instructions "in-flight" at the same time and schedule instructions dynamically (i.e. support out-of-order execution). What advantages does dynamic scheduling offer when compared to an in-order superscalar processor? [4 marks]

(*d*) Some processors convert short-forward branches, i.e. those that branch over a few instructions, to an instruction that sets a predicate register followed by a short sequence of instructions that are conditionally executed depending on the value of the predicate.

(*i*) Why might such a scheme perform better than simply relying on branch prediction and what are its limitations? [4 marks]

(*ii*) Given a superscalar processor that supports out-of-order execution, briefly outline what changes to the processor would be required to support such a scheme. [4 marks]

2 Bioinformatics

(*a*) Calculate the dynamic programming matrix and one or more optimal alignment(s) for the sequences `GAATTC` and `GATTA`, scoring +2 for a match, −1 for a mismatch and with a linear gap penalty of $d = 2$. [5 marks]

(*b*) Determine whether the RNA string `GGACCACCAGG` should be folded into two substructures. [7 marks]

(*c*) Discuss how to carry out the multiple sequence alignment of the following three sequences: `TTTTAAAA`, `AAAACCCC`, `CCCCTTTT`. [4 marks]

(*d*) Discuss which steps of the 1994 Adleman's DNA computation approach would particularly affect the scalability of the number of the visited cities. [4 marks]

3 Cryptography

Your colleagues need a pseudo-random permutation $P_K : \mathbb{Z}_{10^6} \leftrightarrow \mathbb{Z}_{10^6}$, over the integers in the range 0 to 999 999, where K is a 128-bit key. The standard library of their development environment offers them only a 128-bit pseudo-random permutation, in form of the blockcipher AES-128.

(*a*) Recalling that $2^{20} = 1.048576 \times 10^6$, they first decide that implementing a 20-bit pseudo-random permutation $T_K : \{0,1\}^{20} \leftrightarrow \{0,1\}^{20}$ might get them closer to a solution. How could they implement T_K using the available AES_K function? [4 marks]

(*b*) One of your colleagues then proposes to use the function

$$P'_K(m) := \langle T_K(\langle m \rangle_{20}) \rangle^{-1} \bmod 10^6$$

as a "good enough" approximation of what is required.

Notation: $\langle \cdot \rangle_n : \mathbb{Z}_{2^n} \to \{0,1\}^n$ encodes non-negative integers as n-bit bitstrings and $\langle \cdot \rangle^{-1} : \{0,1\}^* \to \mathbb{N}$ does the opposite, i.e. $\langle \langle i \rangle_n \rangle^{-1} = i$ for all $0 \leq i < 2^n$.

Propose a distinguisher D that can distinguish P'_K from a random permutation $R : \mathbb{Z}_{10^6} \leftrightarrow \mathbb{Z}_{10^6}$ using not more than 5000 oracle queries, and show that it achieves $\left|\mathbb{P}(D^{P'_K(\cdot)} = 1) - \mathbb{P}(D^{R(\cdot)} = 1)\right| > \frac{1}{2}$ averaged over all K. [6 marks]

(*c*) Another colleague then proposes the following algorithm:

```
function P_K(m):
  c := T_K(⟨m⟩_20)
  m := ⟨c⟩^-1
  while m ≥ 10^6:
    c := T_K(c)
    m := ⟨c⟩^-1
  return m
```

Show that this is in fact a permutation by

(*i*) explaining why this algorithm always terminates; [1 mark]

(*ii*) providing an implementation of the inverse $P_K^{-1}(m)$. [3 marks]

(*d*) What side-channel risk could the algorithm for $P_K(m)$ from part (*c*) pose, and what can an observer learn from it? [2 marks]

(*e*) Propose an alternative algorithm that reduces the risk that an observer can learn anything from this type of side channel to a negligible probability. [4 marks]

4 Denotational Semantics

Define Σ to be the flat domain $\{ \top \}_\perp$.

(a) Let P be a poset with partial order $\sqsubseteq$ and let S be a subset of P.

Define $\mathcal{M}(S)$ to be the property of S given by $\forall\, x \in S. \forall\, y \in P. x \sqsubseteq y \;\Rightarrow\; y \in S$.

Prove that $\mathcal{M}(S)$ holds if, and only if, there exists a monotone function $f : P \to \Sigma$ such that $f^{-1}\{\top\} = S$. [8 marks]

(b) Let D be a cpo.

(i) For subsets S of D, define a property $\mathcal{C}(S)$ in terms of the cpo structure of D such that

$$(\dagger) \left\{ \begin{array}{l} \mathcal{C}(S) \text{ holds if, and only if, there exists a continuous function} \\ f : D \to \Sigma \text{ such that } f^{-1}\{\top\} = S. \end{array} \right.$$

[4 marks]

(ii) Prove (†) above. [8 marks]

5 E-Commerce

Due to recent changes in the market it is felt that there is an opportunity to build a new social networking platform.

(a) Describe two business models, other than the traditional advertising supported model, along with their implications for the social network in terms of product development and scalability. [6 marks]

(b) What legal considerations should you keep in mind if you were to develop an advertising-based business model? [4 marks]

(c) Outline a marketing plan for an advertising-based social network that will grow the network's revenue to the breakeven point. [10 marks]

6 Hoare Logic and Model Checking

Consider a programming language with commands C consisting of the `skip` no-op command, sequential composition C_1`;`C_2, loops `while` B `do` C for Boolean expressions B, conditionals `if` B `then` C_1 `else` C_2, assigment X `:=` E for program variables X and arithmetic expressions E, heap allocation X `:= alloc(`E_1`,...,`E_n`)`, heap assignment `[`E_1`] :=` E_2, heap dereference X `:= [`E`]`, and heap location disposal `dispose(`E`)`. Assume `null` $= 0$, and predicates for lists and partial lists:

$$\begin{aligned}
&\mathrm{list}(t, []) = (t = \mathrm{null}) \wedge \mathit{emp}\\
&\mathrm{list}(t, h :: \alpha) = \exists y.(t \mapsto h) * ((t+1) \mapsto y) * \mathrm{list}(y, \alpha)\\
&\mathrm{plist}(t_1, [], t_2) = (t_1 = t_2) \wedge \mathit{emp}\\
&\mathrm{plist}(t_1, h :: \alpha, t_2) = \exists y.\ (t_1 \mapsto h) * ((t_1+1) \mapsto y) * \mathrm{plist}(y, \alpha, t_2)
\end{aligned}$$

In the following, all triples are linear separation logic triples.

(a) Explain why a command C of your choice satisfies the following triple, or explain why no such C exists: $\{\texttt{null} \mapsto 5\}\ C\ \{\top\}$. [2 marks]

(b) Explain why a command C of your choice satisfies the following triple (i.e. moves v to a different location): $\{x \mapsto v \wedge X = x\}\ C\ \{Y \mapsto v \wedge Y \neq x\}$. [2 marks]

(c) Give a loop invariant that would serve to prove the following triple, for a command that creates a reversed copy of a list (no proof outline required).
$\{\mathrm{list}(X, \alpha)\}$
`Y := null; C := X;`
`while C` $\neq$ `null do (V := [C]; Y := alloc(V,Y); C := [C+1])`
$\{\mathrm{list}(X, \alpha) * \mathrm{list}(Y, \mathrm{rev}\ \alpha)\}$ [4 marks]

(d) Adjust the program in (c) with a new loop body C_L, so it (still) terminates and $\{\mathrm{list}(X, \alpha)\}$ `Y := null; C := X; while C` $\neq$ `null do` C_L $\{\mathrm{list}(Y, \mathrm{rev}\ \alpha)\}$ holds (no proof, loop invariant, or termination argument required). [2 marks]

(e) Consider an *unsound* extension of the separation-logic proof system with the rule $\{E_1 > 0 \wedge \mathrm{emp}\}$ `alloc_here`(E_1, E_2) $\{E_1 \mapsto E_2\}$ for a new command `alloc_here`(E_1, E_2). Explain in detail, with reference to the proof rules, how $\{\mathrm{emp}\}\ C\ \{\bot\}$ is derivable, for a non-looping C of your choice. [4 marks]

(f) Give a loop invariant that would serve to prove the following triple, for a command that creates a list of the Fibonacci numbers up to n (no proof outline required). Assume $\mathrm{fibs}(i, j) = [\mathrm{fib}\ i, \ldots, \mathrm{fib}\ j]$ for $i \leq j$ and [] otherwise.
$\{\mathrm{emp} \wedge (N = n \wedge n > 2)\}$
`II := alloc(1,null); I := alloc(0,II); X := I; C := 2;`
`while C ≤ N do` (`IV := [I]; IIV := [II]; I := II;`
`II := alloc(IV+IIV,null); [I+1] := II; C := C+1`)
$\{\mathrm{list}(X, \mathrm{fibs}(0, n))\}$ [6 marks]

7 Information Theory

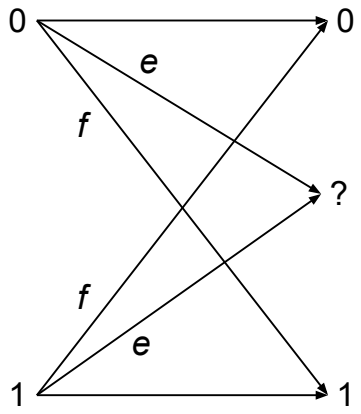
(a) Describe the concepts of discrete entropy and mutual information and how they relate to each other. [4 marks]

(b) How does your answer to part (a) change when the system is continuous rather than discrete? [2 marks]

(c) How do entropy and mutual information relate to the capacity of a noisy channel? [3 marks]

(d) Consider a noisy binary channel with input X and output Y. Under what circumstances is $H(Y|X)$ independent of the distribution of X? [3 marks]

(e) A noisy binary channel is modeled as shown in the diagram below:



The probability of a bit being flipped is f. The probability of a bit being erased is e. Derive the capacity of this channel and the probability distribution of the input bits that achieves it. [8 marks]

You may use the following equality without proof:

$$\begin{aligned} H(a, 1-a-b) &= -a\log_2(a) - (1-a-b)\log_2(1-a-b) \\ &= (1-b)H\left(\frac{a}{1-b}, 1-\frac{a}{1-b}\right) - (1-b)\log_2(1-b) \end{aligned}$$

8 Machine Learning and Bayesian Inference

(*a*) State the *central limit theorem* for a sequence $X_1, X_2, \ldots, X_n$ of independently and identically distributed (iid) random variables having mean $\mathbb{E}(X) = \mu$ and variance $\text{var}(X) = \sigma^2$. [2 marks]

(*b*) Explain how the central limit theorem can be used to provide a *two-sided confidence interval* for the estimate of the mean of a random variable. [4 marks]

(*c*) When $X \in \{0, 1\}$, explain how an estimate of $\mathbb{E}(X)$ can be used to obtain an estimate of $\text{var}(X)$. [2 marks]

(*d*) We know that if X is normal distributed with mean 0 and variance 1 then $\Pr(-1.96 \leq X \leq 1.96) > 0.95$. You have trained a classifier using algorithm A on a data set and tested it using 1000 test examples. You obtain 57 errors. Find a two-sided 95% confidence interval for the accuracy. [3 marks]

(*e*) You have a trained second classifier using algorithm B and the same data set as in Part (*d*), which, when tested using the same 1000 test examples, gives 55 errors. Denoting by a the actual accuracy of the first classifier, and by a' the actual accuracy of the second classifier, find a two-sided 95% confidence interval for the difference $(a - a')$ in the accuracies of the two classifiers. State any new assumptions that you make. [4 marks]

(*f*) Your boss has a vested interest in arguing that algorithm B is better than algorithm A for the problem of interest, and argues that the measured accuracies confirm this view. Discuss the pros and cons of this conclusion, and suggest how you might conduct further experiments to help your boss. [5 marks]

9 Optimising Compilers

The following excerpt from a program in C-style code is optimised by a compiler using data-flow analyses and transformations. Assume that variables `x`, `y` and `z` have already been defined:

```
a = x - y
if (a > 3) {
    b = a + z
    c = x - y
} else {
    b = a + z
    a = a * b
}
d = x - y
b = b / d
print(a * b)
```

(*a*) Using available expression analysis, perform common subexpression elimination on the code showing the results of both the analysis and transformation. [5 marks]

(*b*) Using very busy expression analysis, perform code hoisting on the code from part (*a*) showing the results of both the analysis and transformation. [4 marks]

(*c*) Using reaching definition analysis, perform copy propagation on the code from part (*b*) showing the results of both the analysis and transformation. [*Hint:* use the results of the analysis to transform across basic blocks.] [4 marks]

(*d*) Using live variable analysis, perform dead code elimination on the code from part (*c*) showing the results of both the analysis and transformation. [4 marks]

(*e*) Perform *if* simplification on the code from part (*d*) showing the result of the transformation. [3 marks]

10 Principles of Communications

(*a*) A mobile host moves while transmitting packets via IP multicast. What effect will this have on multicast routing algorithms? [10 marks]

(*b*) A mobile host is seeing a high rate of packet loss. How will this affect the throughput seen by a TCP connection that it has to some fixed server? [10 marks]

11 Quantum Computing

(*a*) What problem does Grover's search algorithm tackle, and what is its advantage over the best classical algorithm for this task? [2 marks]

(*b*) Let there be a database containing 32 elements, indexed by the binary numbers 00000 to 11111. A single element 00110 is *marked*.

(*i*) Give an oracle circuit that identifies the marked element. [1 mark]

(*ii*) If Grover's search algorithm is applied to find the marked element, what should the initial state be set to, and what is the state after a single Grover iterate has been applied? [4 marks]

(*iii*) To find the marked element with maximum probability requires N iterates in total. What is the value of N, and what is the probability of correctly finding the marked element? [4 marks]

(*iv*) If the algorithm is instead run with $3N$ iterates in total, what is the probability of correctly finding the marked element? Comment on your answer. [2 marks]

(*c*) Let V be an oracle circuit that marks one or more elements, acting as follows:

$$V(|x\rangle|a\rangle) = |x\rangle|a \oplus f(x)\rangle$$

Here a takes the values 0 or 1, and we have $f(x) = 1$ when x is the index of a marked element, and $f(x) = 0$ otherwise. How could V be altered to allow Grover's search to find an *unmarked* element? [2 marks]

(*d*) A Grover iterate consists of the oracle circuit, typically denoted V, followed by a circuit W:

(*i*) What is the function of W? [1 mark]

(*ii*) What would happen if the order of V and W were swapped, such that Grover's algorithm is run with V following W as the Grover iterate? [4 marks]

12 Randomised Algorithms

In this question, all considered graphs are undirected and d-regular.

(a) State the definition of conductance. [2 marks]

(b) If G is disconnected, what does this imply in terms of the conductance? [1 mark]

(c) If G is disconnected, what does this imply in terms of the spectrum of $\mathbf{L}$? Briefly justify your claim. [3 marks]

(d) The d-dimensional hypercube with $n = 2^d$ vertices is defined by creating a vertex for each d-digit binary number $(x_1, x_2, \ldots, x_d) \in \{0,1\}^d$. Further, any two vertices are adjacent if and only if their binary representations differ in exactly one digit.

(i) Identifying each binary representation $(x_1, x_2, \ldots, x_d) \in \{0,1\}^d$ with a unique vertex label in $\{1, 2, \ldots, n\}$, verify that $f_{(x_1,x_2,\ldots,x_d)} = (-1)^{x_1}$ is an eigenvector of the Laplacian Matrix $\mathbf{L}$ of the hypercube. State the associated eigenvalue of f for both $\mathbf{L}$ and the adjacency matrix $\mathbf{A}$. [7 marks]

(ii) Apply the Spectral Clustering Algorithm to estimate the conductance of the hypercube, assuming that f in (d)(i) is the eigenvector of λ_2. [*Hint:* It suffices to apply the $(n/2)$-th sweep cut only.] [5 marks]

(iii) Combining (d)(ii) and (d)(i), what can you conclude about the found cut? [2 marks]

13 Types

Consider Gödel's T, the simply-typed lambda calculus with function and natural number types, with zero, successor and iterator term formers for the natural number type.

(*a*) Define a logical relation suitable for establishing the termination of closed programs in this language. [5 marks]

(*b*) State the fundamental lemma for this language. [3 marks]

(*c*) State formally what it means for a set of terms X to be "closed under reduction". [2 marks]

(*d*) Prove the fundamental lemma holds for the iterator case. [10 marks]

END OF PAPER