COMPUTER SCIENCE TRIPOS Part IB

Tuesday 3 June 2014 1.30 to 4.30 pm

COMPUTER SCIENCE Paper 4

*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	SPECIAL REQUIREMENTS
Script paper	*Approved calculator permitted*
Blue cover sheets	
Tags	
Rough work pad	

1 Artificial Intelligence I

Evil Robot's dream of lasting romance remains, alas, just a dream. His latest obsession is a toasted sandwich-maker called SN00005833. In order to win her affections he plans to buy her a chocolate muffin from her favourite shop—*Fat Finbar's World of Cake*—before stealing a bunch of flowers from the local cemetery, gift-wrapping the presents, and presenting both gifts to her. Evil Robot's internal systems have been constructed using the *situation calculus* and a theorem prover.

(*a*) Describe the *situation calculus*, concentrating on the fundamental elements that you would expect to see independently of any specific problem. [5 marks]

(*b*) Suggest *two* logical formulae that might appear in Evil Robot's knowledge base in order to describe the initial state for the above problem. [2 marks]

(*c*) Give *two* examples of a *possibility axiom* that might appear in the knowledge base. [4 marks]

(*d*) Give *two* examples of a *successor-state axiom* that might appear in the knowledge base. One of these should in addition address the *ramification problem.* Explain how it does this. [6 marks]

(*e*) Give *one* example of a *unique names axiom* and *one* example of a *unique actions axiom* that might appear in Evil Robot's knowledge base for this problem. Explain why such axioms are required. [3 marks]

2 Artificial Intelligence I

This question relates to binary *constraint satisfaction problems (CSPs).* A CSP has a set $X = \{x_1, \ldots, x_n\}$ of variables, each having a domain $D_i = \{v_1, \ldots, v_{n_i}\}$ of values. In addition, a CSP has a set $C = \{C_1, \ldots, C_m\}$ of constraints, each relating to a subset of X and specifying the allowable combinations of assignments to the variables in that subset.

(*a*) Give a general definition of a *solution* to a CSP. [1 mark]

(*b*) Given a binary CSP, define what it means for a directed arc $x_i \rightarrow x_j$ between variables x_i and x_j to be *arc consistent.* [2 marks]

(*c*) Give an example of how a directed arc $x_i \rightarrow x_j$ can fail to be arc consistent. Explain how this can be fixed. [2 marks]

(*d*) Describe the *AC-3* algorithm for enforcing arc consistency. [5 marks]

(*e*) Prove that the time complexity of the *AC-3* algorithm is $O(n^2 d^3)$ where d is the size of the largest domain. [3 marks]

(*f*) Suggest a way in which the concept of arc consistency, also known as *2-consistency* can be extended to sets of three, rather than two variables. In the remainder of the question we will refer to this as *3-consistency.* [1 mark]

(*g*) Give an example of how a set of three variables might fail to be 3-consistent, and show how 3-consistency might then be imposed. [2 marks]

(*h*) Suggest a modified version of the *AC-3* algorithm that can be used to enforce 3-consistency. [4 marks]

3 Computer Graphics and Image Processing

Given a sequence of points $(V_i)_{i=0}^{n}$ on a plane, consider the problem of interpolating a smooth curve through all of the points in order by constructing a sequence of polynomial parametric functions, one for each interval $[V_i, V_{i+1}]_{i=0}^{n-1}$.

(*a*) What is meant by C_k continuity at the junction between two curve segments? [2 marks]

(*b*) Explain how the degree of the polynomial function for a curve segment constrains the continuity at its two ends. What continuity can be achieved at each end of a cubic segment? [4 marks]

(*c*) Derive a cubic parametric function for the interval $[V_i, V_{i+1}]$ where $0 < i < n-1$. [10 marks]

(*d*) What special provision would have to be made for the segments $[V_0, V_1]$ and $[V_{n-1}, V_n]$? [4 marks]

4 Computer Graphics and Image Processing

Given a model of a scene represented as a set of triangles in three-dimensional space defining its surfaces, consider the problem of rendering it on a raster display. Write brief notes on:

(*a*) the data that would be stored for each triangle; [2 marks]

(*b*) perspective projection from an arbitrary viewpoint; [5 marks]

(*c*) clipping the data to a suitable viewing frustrum; [5 marks]

(*d*) identifying pixels on the screen within a triangle; [3 marks]

(*e*) resolving hidden surfaces using a z-buffer. [5 marks]

5 Databases

Suppose that an Entity-Relationship model has been constructed that contains two entities $S(\underline{A},\ B)$ and $T(\underline{C},\ \mathit{Amount})$, where A, B, C and *Amount* are attributes and the underline indicates a key. Suppose that we also have a many-to-many relationship R between S and T.

We might expect that this model would be implemented in a relational schema such as $S(\underline{A},\ B)$, $T(\underline{C},\ \mathit{Amount})$, and $R(\underline{A},\ \underline{C})$. However, the database implementor has noticed that a very common and expensive query is this: given an A-value a, find the sum of all *Amount* values for records in T related to this a value in S. Therefore, the implementor has decided to "optimise" the database and replace table S with S' having schema

$$S'(\underline{A},\ B,\ \mathit{Sum}),$$

where the records in table S' will contain the precomputed values for this query. In this way the common and expensive query can be answered by a single key-based read. (Note: *Sum* should be 0 if no matching records exist.)

(*a*) Explain how the operation *insert* $(a,\ b)$ *into* S can be correctly implemented in the $\{S',\ R,\ T\}$ database. [4 marks]

(*b*) Explain how the operation *insert* (c, v) *into* T can be correctly implemented in the $\{S',\ R,\ T\}$ database. [4 marks]

(*c*) Explain how the operation *insert* $(a,\ c)$ *into* R can be correctly implemented in the $\{S',\ R,\ T\}$ database. [4 marks]

(*d*) For an OLTP database, discuss the performance implications of this so-called optimisation. [4 marks]

(*e*) This example illustrates a fundamental trade-off in the design and implementation of database applications. Discuss. [4 marks]

6 Databases

(*a*) We are given a relational schema $R(A,\ B,\ C, D,\ E)$ and told that the following table represents a legal instance of R.

A	B	C	D	E	*tuple number*
1	2	5	4	3	(#1)
1	4	5	4	4	(#2)
2	4	5	4	5	(#3)
2	5	5	4	3	(#4)

Which of the following sets of functional dependencies *may* hold in R? If a set of dependencies cannot hold, then explain why. You can refer to *tuple numbers* in your explanation.

(*i*) F_1 is the set $\{A \to D\}$. [2 marks]

(*ii*) F_2 is the set

$$\begin{aligned} A, B &\to C \\ E &\to B \\ D, E &\to A \end{aligned}$$

[2 marks]

(*iii*) F_3 is the set

$$\begin{aligned} A, B &\to C \\ D, E &\to C \\ A &\to D \end{aligned}$$

[4 marks]

(*b*) We are given a relational schema $R(\mathbf{Z},\ \mathbf{W},\ \mathbf{Y})$. Suppose that in some (correct) instance of R the query

$$(\pi_{\mathbf{Z},\mathbf{W}}(R) \bowtie \pi_{\mathbf{Z},\mathbf{Y}}(R)) - R$$

is not empty. What can we conclude about the functional dependency $\mathbf{Z} \to \mathbf{W}$? Explain your answer. [4 marks]

(*c*) In the process of using functional dependencies to normalise a schema, what is meant by a *lossless join decomposition* and how is such a decomposition guaranteed? [4 marks]

(*d*) In schema normalisation, is Boyce-Codd Normal Form (BCNF) always to be preferred over 3rd Normal Form (3NF)? Explain your answer. [4 marks]

7 Economics, Law and Ethics

(*a*) Describe the provisions of the Data Protection Act. [8 marks]

(*b*) You are designing and are about to launch a mobile phone app which will seek to understand the emotional condition of the user, using multiple inputs such as motion sensing, facial expression recognition, voice stress measurement and the analysis of entered text. Its declared purpose is to enable services to interact more empathically with users. You propose to monetize it by serving ads at times when the user is more likely to buy. Your "backers" have raised a concern that this app will be able to diagnose depression, and that in consequence you may be storing substantial amounts of sensitive personal information.

Discuss this problem from the viewpoints of both data protection law and ethics. [12 marks]

8 Security I

(*a*) Windows implements *static inheritance* for the access-control lists of NTFS files and folders.

(*i*) What does *static inheritance* mean here and how does it differ from *dynamic inheritance*? [4 marks]

(*ii*) Five flag bits (`ci`,`oi`,`np`,`io`,`i`) in each NTFS access-control entry (ACE) manage how it is inherited. Briefly describe the purpose of each bit. [5 marks]

(*iii*) User `mike` gives his folder `project` the following access-control list:

```
project
   AllowAccess mike: full-access (oi,ci)
   AllowAccess alice: read-execute (ci,np)
   AllowAccess bob: read-only (oi)
```

It contains one folder and two text files, none of which have any non-inherited access-control entries:

```
project\doc.txt
project\src
project\src\main.c
```

For each of these three objects, list all inherited access-control entries, showing in parentheses the inheritance-control flag bits that are set (using the same notation as above). [5 marks]

(*b*) Describe the purpose and four typical functions of a *root kit.* [6 marks]

9 Security I

Block ciphers usually process 64 or 128-bit blocks at a time. To illustrate how their modes of operation work, we can use instead a pseudo-random permutation that operates on the 26 letters of the English alphabet:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
m	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
$E_K(m)$	P	K	X	C	Y	W	R	S	E	J	U	D	G	O	Z	A	T	N	M	V	F	H	L	I	B	Q

As the XOR operation is not defined on the set $\{\mathtt{A}, \dots, \mathtt{Z}\}$, we replace it here during encryption with modulo-26 addition (e.g., $\mathtt{C} \oplus \mathtt{D} = \mathtt{F}$ and $\mathtt{Y} \oplus \mathtt{C} = \mathtt{A}$).

(*a*) Encrypt the plaintext "TRIPOS" using:

(*i*) electronic codebook mode; [2 marks]

(*ii*) cipher-block chaining (using IV $c_0 = \mathtt{K}$); [4 marks]

(*iii*) output feedback mode (using IV $c_0 = \mathtt{K}$). [4 marks]

(*b*) Decrypt the ciphertext "BSMILVO" using cipher-block chaining. What operation should replace XOR? [4 marks]

(*c*) Your opponent is allowed to send you two plaintext messages M_0 and M_1, each n letters long. You now pick a new private key K, resulting in a new pseudo-random permutation $E_K : \{\mathtt{A}, \dots, \mathtt{Z}\} \leftrightarrow \{\mathtt{A}, \dots, \mathtt{Z}\}$. You also pick uniformly at random a private bit $b \in \{0, 1\}$ and return a ciphertext $C = c_0 c_1 \dots c_n$, namely the message M_b encrypted with cipher-block chaining using the fresh E_K. Finally, your opponent has to guess your bit b.

Approximately how large must n be at least for your opponent to have a greater than 75% chance of guessing b correctly? Outline a strategy that your opponent can use to achieve this. [6 marks]

END OF PAPER