COMPUTER SCIENCE TRIPOS Part II

Tuesday 3 June 2008 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 Additional Topics

The following protocol is meant to establish a strong shared secret between two wireless devices A and B through a Diffie–Hellman exchange over radio. To guard against man-in-the-middle attacks, in message 3 device A sends device B a 16-bit secret random value R over a different channel, for example by showing the value on A's screen and having the human user retype it into B's keypad.

Notation: $x|y$ indicates the concatenation of bit strings x and y, while $m_K(x)$ indicates the MAC (message authentication code) of message x using key K.

(1)	$A \rightarrow B$	:	g^a
(2)	$A \leftarrow B$	:	g^b
(3)	$A \rightarrow B$	:	R
(4)	$A \rightarrow B$	:	$m_{K_A}(A\|g^a\|g^b\|R)$
(5)	$A \leftarrow B$	:	$m_{K_B}(B\|g^a\|g^b\|R)$
(6)	$A \rightarrow B$	:	K_A
(7)	$A \leftarrow B$	:	K_B
(8)	(on their own)	:	(verification)
(9)	$A \leftrightarrow B$	:	(confirmation)

(a) Explain what the resulting shared secret will be and what additional verification and confirmation steps each side must take after exchanging the first 7 messages shown above. [3 marks]

In the following questions, "explain *in detail*" means with reference to the exact messages exchanged and expected by A, B and a man-in-the-middle M; and, where appropriate, with suitable protocol diagrams involving all three.

(b) Explain in detail how a man-in-the-middle M could successfully attack this protocol if R were not used or if M could eavesdrop on message 3. [4 marks]

(c) Explain in detail how the introduction of R stops the man-in-the-middle. [5 marks]

(d) Explain in detail how the man-in-the-middle could still successfully attack this protocol if the confirmation of step 9 were omitted. [8 marks]

2 Additional Topics

(*a*) Give a brief description of the three segments that make up the Global Positioning System (GPS), and their main software, hardware and information components. [5 marks]

(*b*) What are the main timing observables from the GPS signals, and how can they be processed to give time, position and velocity? [5 marks]

(*c*) What data is broadcast in the GPS signals, and how is it used by receivers? [5 marks]

(*d*) Explain some of the ways in which an additional communication channel can be used to assist a GPS receiver, and what advantages the user might observe. [5 marks]

3 Digital Communication II

(*a*) Describe the OSI (Open Systems Interconnections) reference model and discuss examples of the networking functions normally associated with each component. [14 marks]

(*b*) What do we mean by *layer violation*? Discuss some of the reasons that might lead a pragmatic protocol implementer to engage in layer violation and illustrate with an example. [6 marks]

4 Distributed Systems

(*a*) You have been asked to design an event composition and aggregation service to operate above publish/subscribe middleware to be deployed in various environmental monitoring scenarios. Your service is to advertise and publish high-level events of interest to applications. It may subscribe to any published events in the domain of deployment. Your service may itself be distributed.

Identify a list of technical design issues you would raise with your client prior to specifying the service in detail. [12 marks]

(*b*) By means of a diagram illustrate the use of vector clocks to implement the delivery in causal order of multicast messages among members of a closed, unstructured process group. Include an explanation of the message delivery algorithm. [8 marks]

5 Comparative Architectures

(*a*) Why are multi-level caches often used in preference to a single larger cache? How might the parameters of an L1 and L2 data cache typically differ? [8 marks]

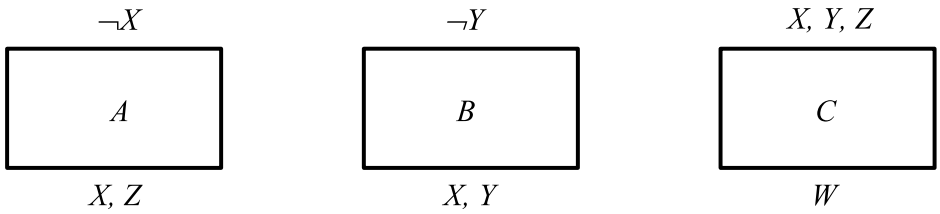
(*b*) What does it mean if a cache memory hierarchy adopts a multi-level inclusion policy? What might influence a decision on whether to adopt a multi-level inclusion or exclusion policy? [6 marks]

(*c*) A processor's multi-level cache hierarchy consists of L1 and L2 caches with the following characteristics: L1 miss rate is 2%, L1 hit time is 2 cycles, local L2 miss rate is 20%, L2 hit time is 10 cycles, L2 miss penalty is 200 cycles. It is suggested that reducing the size of the L1 cache will improve overall performance by reducing the L1 cache's hit time to a single cycle. The reduction in L1 cache size will increase the L1's miss rate to 3%. Will average memory access time actually be improved? Clearly state any formulae you use and show your calculations. [6 marks]

6 Artificial Intelligence II

We have a simple, propositionalised planning problem and we suspect that we might be able to solve it using the *GraphPlan* algorithm. The problem is as follows.

Action A has preconditions $\{\neg X\}$ and effects $\{X, Z\}$, action B has preconditions $\{\neg Y\}$ and effects $\{X, Y\}$, and action C has preconditions $\{X, Y, Z\}$ and effects $\{W\}$. The start state for the problem is $\{\neg W, \neg X, \neg Y, \neg Z\}$ and the goal is $\{W\}$.



(*a*) Labelling the start state as level S_0 and the first action level as A_0, draw the planning graph for this problem up to and including level S_2. Use an entire sheet of paper for this diagram. [5 marks]

(*b*) Describe each of the five kinds of *mutex link* that can appear in a planning graph, and add an example of each to the diagram drawn in part (*a*), clearly labelling it to show which kind of mutex link it is. [10 marks]

(*c*) What is the *level cost* of a literal in a planning graph? Explain why this measure of cost might perform poorly as a measure of how hard the literal is to achieve. [2 marks]

(*d*) Will GraphPlan be able to extract a working plan from the diagram you have drawn in parts (*a*) and (*b*)? Explain your answer. You may if you wish add further mutex links to your diagram at this stage. [3 marks]

7 Advanced Systems Topics

A computer system provides a compare-and-swap (CAS) operation which is used in the following manner:

```
seen = CAS (address, old, new)
```

It loads the contents of `address`, compares the value against `old` and if it matches stores the value `new` at the same address. All of this is performed atomically and the value read from the address is returned as `seen`.

(*a*) What does it mean for a processor instruction to be *atomic*? [2 marks]

(*b*) Write pseudocode for a simple spin lock using `CAS`. [4 marks]

Consider a singly-linked list of `QNode` objects, each with a Boolean field `value` and a reference `next` to its successor (holding `null` at the tail of the queue). A shared location `l` refers to the tail node (or is `null` if the queue is empty).

(*c*) Define the following concurrent operations using `CAS`:

```
// Append a new node q to the tail of the list, returning
// the previous tail
QNode pushTail (QNode q);

// Remove q, the current head of the list, returning
// the new head
QNode popHead (QNode q);
```

[8 marks]

(*d*) Define a *queue-based spin lock* based on these operations. [6 marks]

8 Advanced Systems Topics

(*a*) Explain the difference between *distance vector* routing and *link-state* routing. [2 marks]

(*b*) Describe in detail the problem of *counting to infinity* and describe two solutions. [4 marks]

(*c*) Compare and contrast the two existing methods of damping route instability in the Border Gateway Protocol (BGP). [6 marks]

(*d*) Could link-state protocols be used for interdomain routing? Explain your answer. [8 marks]

9 Security

(*a*) What does it mean for a hash function to be *collision-resistant*, and to be *preimage-resistant*? [4 marks]

(*b*) The current scheme for doing background checks on schoolteachers, health service staff and others who have contact with children is too slow, and your job is to design its replacement. You are given a database of 20,000 convicted sex offenders stored as (date of birth, name). You may not release any information that might identify an offender. You may only release signed information providing evidence that a supplied input of (date of birth, name) does not appear on the offenders' database. Finally, because there are huge peaks in transaction volume at the start of the school year and when National Health Service staff rotate jobs, you want all – or almost all – digital signatures to be precomputed for performance reasons.

Provide an outline design for the system and show how it meets the requirements. [16 marks]

10 Natural Language Processing

(*a*) In the context of evaluation of NLP systems, what is meant by the terms *baseline* and *ceiling*? [3 marks]

(*b*) Discuss the evaluation of:

(*i*) stochastic part of speech (POS) tagging; [4 marks]

(*ii*) word sense disambiguation; [4 marks]

(*iii*) pronoun resolution. [4 marks]

(*c*) What additional considerations may apply when the evaluation concerns a component of a complete NLP application? [5 marks]

11 Information Retrieval

Question Answering, i.e. the automatic creation of answers to factual natural language questions, is a relatively recent research topic in NLP/IR. It is evaluated in competitive evaluations, e.g., the TREC-QA competitions.

(*a*) Three different evaluation metrics have been used in TREC-QA. Define these three metrics, specify in which form the metrics require the system output, and illustrate them using the example questions and answers given in Figure 1 on the next page. [7 marks]

(*b*) There are factual questions that even humans find hard to answer, and such questions are harder for systems to answer as well. Name at least *three* types of such questions, and argue why they are difficult. [6 marks]

(*c*) There are ways in which participants can "cheat" in TREC-QA ("cheating" is defined as "changing one's behaviour in such a way as to get higher scores, without a real improvement in performance").

(*i*) Describe *two* different ways of cheating.

(*ii*) What countermeasures could the organisers of TREC-QA take to protect themselves against these ways of cheating?

[7 marks]

Question 1: When was Volkswagen founded?

By the early 1990s, Volkswagen's annual sales in the US were below...
VW was founded in 1937 as a public concern by the then Nati...
New wage-model at VW, Germany, 2004
factory in Russia, VW now also founded a new company for that purpose.
Founded in 1955, Volkswagen of America, Inc. is headquartered at Auburn...
Volkswagen was founded in 1937 by the German Auto Association.
Oldest Volkswagen Diesel found in California

Question 2: How many calories in a Big Mac?

Big Mac – Calories: 540 (28% USRDA)
Calories in a Big Mac (215g). 429. 23.0. Calories in a Cheeseburger...
Many associate Mcdonald's and other fast foods with high calories and...
How many calories does the average Big Mac from McDonald's restaurants...
NIL
789
540 calories.

Question 3: Who invented the paper clip?

electronics, science and mathematics, invented the paper clip in 1899.
Who invented the paper clip? Ben T.
So, who invented the paper clip? When Johan Vaaler patented his...
Who invented the paper clip and did he or she think it was going to be...
He was employed by the owner of an invention office when he invented ...
Johan Vaaler, the Norwegian Jew who invented the paper clip...

Question 4: When did Madonna die?

NIL
June 2000

Figure 1. Some questions and a system's guess at answers. Boldfaced answers are judged correct.

12 Human–Computer Interaction

You have identified a market opportunity for home media players that would cater for older members of the population. Many older people have difficulty understanding the operating principles of devices such as MP3 players, "internet radios" for streaming audio, and personal video recorders and players.

Describe design and evaluation processes that could be used by a start-up company to improve the usability of such devices for this population.

You should consider several different stages of the product design cycle, and describe *five* different user interface design techniques that would be relevant at those different stages. [4 marks each]

13 Business Studies

(*a*) Describe *five* criteria that an investor might use to evaluate a business. Which is the most important, and why? [5 marks]

(*b*) Distinguish between *marketing* and *selling*. [5 marks]

(*c*) Distinguish between *quantitative* and *qualitative* market research. [5 marks]

(*d*) A PC manufacturer obtains the following results from test marketing PC systems:

System including	Price ex VAT	Number sold
19 inch monitor and software bundle	£299	1000
19 inch monitor, software bundle and a printer	£399	750
22 inch monitor and software bundle	£499	400
22 inch monitor, printer but no additional software	£599	500

If the test market area represents 1% of the target population, what price point and how many sales should be expected for a system with a 22 inch monitor but with neither software nor printer? [5 marks]

14 E-Commerce

(*a*) When designing an interactive web site, describe *five* desirable stylistic points. [5 marks]

(*b*) Discuss the advantages and disadvantages of showing stock levels on a commercial site. [5 marks]

(*c*) A certain UK retailer proposes to sell camera lenses online via a web site. Identify *two* major regulatory regimes applicable, and give examples of the steps required to comply with them. [10 marks]

15 Specification and Verification I

(*a*) Describe how Floyd–Hoare logic can be used to reason about programs that use arrays. [4 marks]

(*b*) Write a partial correctness specification of the form $\{P\}\ C_{max}\ \{Q\}$ that specifies that the effect of executing C_{max} is to set the variable `M` to the maximum of the values `A(0)`, ..., `A(N)` (where `0` $\leq$ `N`) in the state before executing the command (i.e. C_{max} computes the maximum value stored in the array `A` between positions `0` and `N`). [4 marks]

(*c*) Devise a particular command C_{max} that meets your specification. [4 marks]

(*d*) Give an outline proof that your command meets your specification. [8 marks]

16 Specification and Verification II

(*a*) Specify a combinational device `MAX` with two 4-bit inputs i_1, i_2 and a 4-bit output o, such that the value output on o is the input that has the greater value when interpreted as a binary number. [2 marks]

(*b*) Specify a sequential device `REG`(w) with a 4-bit input i and output o such that on the first cycle (cycle 0) w is output on o and on cycle n ($n > 0$) the value input on the preceding cycle is output. [2 marks]

(*c*) Write a specification of a device D_{max} with a 4-bit input i and a 4-bit output o such that the value output on o on the n-th cycle is the maximum value input on i on all cycles up to and including the n-th cycle. [4 marks]

(*d*) Devise a circuit built out of `MAX` and `REG` that implements your specification. [4 marks]

(*e*) Outline how to prove that your circuit meets your specification. [8 marks]

END OF PAPER