CST1
COMPUTER SCIENCE TRIPOS Part IB

Tuesday 10 June 2025 13:30 to 16:30

COMPUTER SCIENCE Paper 5

*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 Computer Networking

Analyze the fundamental aspects of error control in digital communication systems.

(*a*) Describe, compare, and contrast Forward Error Correction (FEC) and Error Detection and Retransmission (ED&R) in terms of their bandwidth utilisation and delay characteristics. [6 marks]

(*b*) A streaming video service needs to implement error control mechanisms. Explain how the channel properties, latency, reliability, and available bandwidth, would influence the choice between FEC and ED&R. [8 marks]

(*c*) (*i*) Consider a CRC polynomial used for error detection: outline the design capability, implementation, and limitations. [3 marks]

(*ii*) Compare and contrast two retransmission strategies for a system implementing ED&R. [3 marks]

2 Computer Networking

Consider the principles of flow and congestion control in network protocols.

(*a*) Describe how window-based congestion control differs from rate-based congestion control in terms of buffer management and throughput optimization. [4 marks]

(*b*) Consider a communication link with propagation delay τ and bandwidth b. For a packet size p, derive expressions for the maximum achievable throughput under:

(*i*) A simple stop-and-wait protocol [4 marks]

(*ii*) A window protocol with window size 2 [4 marks]

(*c*) Explain how buffer overflow conditions are managed in:

(*i*) On-off flow control [4 marks]

(*ii*) Window-based congestion control with Selective Acknowledgments (SACK) [4 marks]

3 Computer Networking

(*a*) Explain hierarchical and flat addressing schemes, including a comparison of each scheme in terms of table size and address space efficiency. [6 marks]

(*b*) Analyze how the following choices affect route selection:

(*i*) Choice of algorithm: Link state versus distance vector approaches [8 marks]

(*ii*) Operational choice: Link cost metrics [6 marks]

4 Concurrent and Distributed Systems

You have been hired to create a supervision room booking system for Old Hall College, Cambridge. The system should allow supervisors to reserve rooms for particular time slots, and it needs to ensure that there are no two bookings for overlapping time periods for the same room, even when desperate supervisors are frantically and concurrently trying to book the last available room.

To ensure fault tolerance the system is required to have multiple replicas. It must work in a partially synchronous system model with crash-recovery faults.

(*a*) Your first thought is to represent each hourly time slot for each supervision room as a separate register, with the value of the register indicating the supervisor who has booked it (if any), and to use the algorithm by Attiya, Bar-Noy, and Dolev (ABD algorithm) to replicate those registers. Assume that every time slot begins on the hour and lasts for an hour, so there are no overlapping slots. Explain briefly how the ABD algorithm works, and which consistency properties such a room booking system would have. Pseudocode is not needed. [6 marks]

(*b*) Your second idea is to instead use state machine replication (SMR). Give pseudocode for a SMR-based supervision room database, and explain the pros and cons of this approach compared to using ABD. You may assume that you already have an implementation of total order broadcast. [6 marks]

(*c*) Your third idea is to use two-phase commit (2PC) to coordinate writes to the replicas. Explain how 2PC works, and how it compares to the previous two approaches in the context of the room booking system. [6 marks]

(*d*) Which of the three approaches (ABD, SMR, 2PC) is best for this system? Briefly explain why. [2 marks]

5 Concurrent and Distributed Systems

You are developing a service that sets up video calls between random pairs of strangers. It works as follows: a user connects to a server when they want to talk to someone. If another user is already connected to the same server and not already talking to someone else, the server connects the two users to each other. If nobody else on the same server is free, the user waits for the next user to connect.

(a) Assume the server uses one thread per connected user to handle that user's connection. Outline (in words, not code) how you might implement a module that determines which users to connect to each other. How will you ensure it is thread-safe when multiple users connect concurrently? [4 marks]

(b) Using Java's monitor API (`synchronized`, `wait()`, `notify()`, `notifyAll()`), write pseudocode for a class with a thread-safe method $\mathsf{matchUser}(u_1) \to u_2$ that takes a user object u_1 and returns another user object u_2. The return value u_2 must be the argument passed by another thread to the same method in a call $\mathsf{matchUser}(u_2) \to u_1$, and that call must return u_1. In other words, u_1 and u_2 are two users whose video feeds will be linked together. You do not need to write any pseudocode for the video handling. If no other user object is currently available for matching, the method must block until another thread has called $\mathsf{matchUser}()$.

Briefly justify why your solution is correct, for example using comments in your pseudocode. Details of syntax and Java APIs are not important; the pseudocode just needs to be clear. [10 marks]

(c) Java offers signal-and-continue semantics for its monitors. Explain what this means, and how you ensure that the code you wrote for Part (b) is correct under this semantics. Also, justify why you used `notify()` or `notifyAll()` in your answer to Part (b). [6 marks]

6 Introduction to Computer Architecture

(*a*) You have designed a simple processor together with private instruction and data caches. You are asked to improve its performance for a particular application. You try each of the changes listed below individually but in each case the performance does not improve significantly. Describe why this might occur.

(*i*) You change the Instruction Set Architecture (ISA) and processor design to provide more registers for the compiler to use. Each instruction is now encoded using more bits. [4 marks]

(*ii*) You add a custom instruction that performs a sequence of instructions in fewer clock cycles than was previously possible. [4 marks]

(*iii*) You replace your current DRAM with a special low-latency DRAM. [3 marks]

(*iv*) You increase the associativity of the data cache from 2-way to 4-way. [3 marks]

(*b*) You add a second processor core on the same chip with its own private caches, but the speedup you achieve for a new multithreaded application is only 1.2×. What different factors could explain this limited speedup? [6 marks]

7 Introduction to Computer Architecture

(*a*) Why might having a formal description of an Instruction-Set Architecture (ISA), for example, for the RISC-V ISA, be useful? [4 marks]

(*b*) In order to achieve the benefits of pipelining it is necessary to balance the delay of logic in different stages. What might make this difficult or expensive to achieve? [4 marks]

(*c*) Why are FPGAs not commonly used for high-volume markets? [4 marks]

(*d*) We run timing analysis after logic synthesis and a particular critical path is reported. We now place and route our design and a different critical path is identified. Why might this happen? [4 marks]

(*e*) After manufacturing our design, we discover that it experiences significant clock skew, meaning the clock signal reaches different flip-flops at different times due to an issue in the clock tree's design. What are the two potential problems this skew could cause? In each case, describe any steps that can be taken to mitigate the issues in the *manufactured* chips even if they result in reduced performance. [4 marks]

8 Introduction to Computer Architecture

(*a*) Why might high-performance system-on-chip (SoC) designs, for example, those that power mobile phones, typically contain many processors, and domain-specific hardware accelerators? [6 marks]

(*b*) If we simply wanted to run independent programs on the different cores of a multicore processor, would we need a cache coherency mechanism, a memory consistency model, both, or neither? [4 marks]

(*c*) What does the term "side-channel" refer to in the context of hardware security? Provide an example of a side channel that could be present in a processor. [4 marks]

(*d*) Why do GPUs typically require high-bandwidth access to main memory? [3 marks]

(*e*) Why do we need to keep Translation Lookaside Buffers (TLBs) coherent in a multi-core system? [3 marks]

END OF PAPER